目　　录

1 基础知识

1.1 概率

概率论的一个基本概念是随机试验:其结果在事先不能确定的试验。所有可能的试验结果的集合称为试验的样本空间,我们以 S 表示。

一个事件是样本空间的一个子集,如果试验的结果是该子集中的元素,则称该事件发生。假设对样本空间 S 的每一个事件 E 定义了一个数 $P(E)$,满足以下三条公理*):

公理(1) $0 \leqslant P(E) \leqslant 1$。

公理(2) $P(S)=1$。

公理(3) 对任何一列互不相容的事件 $E_1, E_2, \cdots$,即 $E_iE_j=\phi,\ i \neq j$ (ϕ 是空集),有 $P(\bigcup_{i=1}^{\infty} E_i)=\sum_{i=1}^{\infty} P(E_i)$。我们称 $P(E)$ 为事件 E 的概率。

公理(1)、(2)及(3)的一些简单推论如下:

1.1.1 如果 $E \subset F$,则 $P(E) \leqslant P(F)$。

*) 事实上 $P(E)$ 只对 S 的所谓可测事件有定义,但我们不涉及这一限制。

1.1.2 $P(E^c)=1-P(E)$，其中 E^c 是 E 的余集。

1.1.3 $P(\bigcup_{i=1}^{n} E_i)=\sum_{i=1}^{n} P(E_i)$，当 E_i 互不相容时。

1.1.4 $P(\bigcup_{i=1}^{\infty} E_i)\leqslant\sum_{i=1}^{\infty} P(E_i)$。

不等式(1.1.4)称为布尔不等式。

概率函数 P 的一个重要性质是连续性。为了更精确地阐明这一性质，需要引进极限事件的概念，其定义如下：若 $E_n\subset E_{n+1}$，$n\geqslant1$，称事件序列$\{E_n,n\geqslant1\}$为递增的；而当 $E_n\supset E_{n+1}$，$n\geqslant1$，时则称为递减的。如果$\{E_n,n\geqslant1\}$是一递增的事件序列，那么我们定义一个新的事件，记为 $\lim\limits_{n\to\infty}E_n$：

$$\lim_{n\to\infty}E_n=\bigcup_{i=1}^{\infty}E_i，\quad 当\ E_n\subset E_{n+1}，\quad n\geqslant1$$

类似地，如果$\{E_n,n\geqslant1\}$是一递减序列，那么定义 $\lim\limits_{n\to\infty}E_n$ 为

$$\lim_{n\to\infty}E_n=\bigcap_{i=1}^{\infty}E_i，\quad 当\ E_n\supset E_{n+1}，\quad n\geqslant1$$

现在我们可以叙述下面的命题：

命题 1.1.1

如果$\{E_n,n\geqslant1\}$是递增的或递减的事件序列，则

$$\lim_{n\to\infty}P(E_n)=P(\lim_{n\to\infty}E_n)$$

证明 首先假设$\{E_n,n\geqslant1\}$是递增序列，并定义事件 F_n，$n\geqslant1$，为

$$F_1=E_1,$$

$$F_n=E_n(\bigcup_{i=1}^{n-1}E_i)^c=E_nE_{n-1}^c,\qquad n>1$$

即 F_n 由包含在 E_n 中但不在任何前面的 E_i $(i<n)$ 中的点组成。容易验证 F_n 是互不相交的事件，满足

$$\bigcup_{i=1}^{\infty}F_i=\bigcup_{i=1}^{\infty}E_i \quad 和 \quad \bigcup_{i=1}^{n}F_i=\bigcup_{i=1}^{n}E_i,\qquad 对一切\ n\geqslant1$$

于是

$$P(\bigcup_{i=1}^{\infty}E_i)=P(\bigcup_{i=1}^{\infty}F_i)$$

$$= \sum_{i=1}^{\infty} P(F_i) \qquad \text{(由公理(3))}$$
$$= \lim_{n\to\infty} \sum_{i=1}^{n} P(F_i)$$
$$= \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} F_i\right)$$
$$= \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} E_i\right)$$
$$= \lim_{n\to\infty} P(E_n)$$

这证明了$\{E_n, n\geqslant 1\}$递增时的结论。

如果$\{E_n, n\geqslant 1\}$是递减序列，则$\{E_n^c, n\geqslant 1\}$是递增序列；因此，

$$P\left(\bigcup_{n=1}^{\infty} E_n^c\right) = \lim_{n\to\infty} P(E_n^c)$$

但因$\bigcup_{n=1}^{\infty} E_n^c = \left(\bigcap_{n=1}^{\infty} E_n\right)^c$，可见

$$1 - P\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} [1 - P(E_n)]$$

或等价地

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} P(E_n)$$

结果得证。

例 1.1 (a)　考虑一个群体，它由能产生同类后代的个体构成。初始的个体数用 X_0 表示，称为第 0 代的总数。第 0 代的全体后代构成第 1 代，其总数以 X_1 表示。一般地，以 X_n 表示第 n 代的总数。

由于 $X_n=0$ 蕴含 $X_{n+1}=0$，所以 $P(X_n=0)$ 是递增的，故$\lim_{n\to\infty} P(X_n=0)$存在。它表示什么呢？要回答这个问题，运用命题 1.1.1 如下：

$$\lim_{n\to\infty} P\{X_n=0\} = P\{\lim_{n\to\infty}\{X_n=0\}\}$$
$$= P\left\{\bigcup_{n=1}^{\infty}\{X_n=0\}\right\}$$
$$= P\{\text{群体迟早灭绝}\}$$

即第 n 代没有个体的极限概率等于此群体最终灭绝的概率。

命题 1.1.1 还能用来证明波莱尔—坎泰利引理。

命题 1.1.2　波莱尔—坎泰利引理

设 E_1，E_2，…为一列事件。如果

$$\sum_{i=1}^{\infty} P(E_i) < \infty$$

则

$$P\{\text{无穷多个 } E_i \text{ 发生}\}=0$$

证明 无穷多个 E_i 发生，这一事件记作 $\limsup_{i\to\infty} E_i$，它可表示为

$$\limsup_{i\to\infty} E_i=\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}E_i$$

因为若有无穷多个 E_i 发生，则 $\bigcup_{i=n}^{\infty}E_i$ 对每个 n 都发生，从而 $\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}E_i$ 发生。另一方面，如果 $\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}E_i$ 发生，则对每个 n，$\bigcup_{i=n}^{\infty}E_i$ 发生，从而对每个 n 至少有一个 $i\geqslant n$ 使 E_i 发生；因此有无穷多个 E_i 发生。

因 $\bigcup_{i=n}^{\infty}E_i$，$n\geqslant 1$，是递减的事件序列，由命题 1.1.1 得

$$\begin{aligned}P\Big(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}E_i\Big)&=P\Big(\lim_{n\to\infty}\bigcup_{i=n}^{\infty}E_i\Big)\\&=\lim_{n\to\infty}P\Big(\bigcup_{i=n}^{\infty}E_i\Big)\\&\leqslant\lim_{n\to\infty}\sum_{i=n}^{\infty}P(E_i)\\&=0\end{aligned}$$

结论得证。

例 1.1(b) 设 X_1，X_2，…使得

$$P\{X_n=0\}=\frac{1}{n^2}=1-P\{X_n=1\},\quad n\geqslant 1$$

如果记 $E_n=\{X_n=0\}$，则因 $\sum_{n=1}^{\infty}P(E_n)<\infty$，由波莱尔—坎泰利引理得，无穷多个 X_n 等于 0 的概率为 0。因此对充分大的 n，X_n 必须等于 1，从而我们可以断定以概率 1 有

$$\lim_{n\to\infty}X_n=1$$

波莱尔—坎泰利引理的逆命题要有独立性的条件。

命题 1.1.3 波莱尔—坎泰利引理的逆命题

设 E_1，E_2，…为独立事件，使得

$$\sum_{n=1}^{\infty}P(E_n)=\infty$$

则

$$P\{\text{无穷多个 } E_n \text{ 发生}\}=1$$

证明

$$
\begin{aligned}
P\{\text{无穷多个 } E_n \text{ 发生}\} &= P\{\lim_{n\to\infty}\bigcup_{i=n}^{\infty} E_i\} \\
&= \lim_{n\to\infty} P(\bigcup_{i=n}^{\infty} E_i) \\
&= \lim_{n\to\infty} [1 - P(\bigcap_{i=n}^{\infty} E_i^c)]
\end{aligned}
$$

现因

$$
\begin{aligned}
P(\bigcap_{i=n}^{\infty} E_i^c) &= \prod_{i=n}^{\infty} P(E_i^c) \qquad \text{(由独立性)} \\
&= \prod_{i=n}^{\infty} [1 - P(E_i)] \\
&\leqslant \prod_{i=n}^{\infty} e^{-P(E_i)} \qquad \text{(由不等式 } 1-x \leqslant e^{-x}\text{)} \\
&= \exp\left(-\sum_{i=n}^{\infty} P(E_i)\right) \\
&= 0, \quad \text{因为对一切 } n, \ \sum_{i=n}^{\infty} P(E_i) = \infty
\end{aligned}
$$

由此结论得证。

例 1.1(c) 设 $X_1, X_2, \cdots$ 独立且使得

$$P\{X_n=0\} = 1/n = 1 - P\{X_n=1\}, \quad n \geqslant 1$$

若记 $E_n = \{X_n = 0\}$，则因 $\sum_{n=1}^{\infty} P(E_n) = \infty$，由命题 1.1.3，无穷多个 E_n 发生。也因 $\sum_{n=1}^{\infty} P(E_n^c) = \infty$，也有无穷多个 E_n^c 发生。因此以概率 1 有无穷多个 X_n 等于 0 及无穷多个 X_n 等于 1，即以概率 1，$n\to\infty$ 时，X_n 不存在极限。

1.2 随机变量

考虑一个随机试验，其样本空间为 S。一个随机变量 X 是一个函数，它给 S 中的每一个结果指定一个实数。对任何实数集 A，X 取的值含于集 A 中的概率等于试验结果包含在 $X^{-1}(A)$ 中的概率。即

$$P\{X \in A\} = P(X^{-1}(A))$$

其中$X^{-1}(A)$是由一切满足$X(s)\in A$的点$s\in S$构成的事件。

随机变量X的分布函数F由下式定义：对任一实数x

$$F(x)=P\{X\leqslant x\}=P\{X\in(-\infty,x]\}$$

我们将以$\overline{F}(x)$记$1-F(x)$，因而

$$\overline{F}(x)=P\{X>x\}$$

一个随机变量X称为离散的，如果它可能取的值的集合是可数的。对于离散随机变量有

$$F(x)=\sum_{y\leqslant x}P\{X=y\}$$

一个随机变量称为连续的，如果存在一函数$f(x)$，称为概率密度函数，使

$$P\{X\text{ 取值于 }B\}=\int_B f(x)dx$$

对每一个集B成立。因为$F(x)=\int_{-\infty}^{x}f(x)dx$，得

$$f(x)=\frac{d}{dx}F(x)$$

两个随机变量X与Y的联合分布函数F定义为

$$F(x,y)=P\{X\leqslant x,Y\leqslant y\}$$

X与Y的分布函数

$$F_X(x)=P\{X\leqslant x\}\quad\text{及}\quad F_Y(x)=P\{Y\leqslant x\}$$

可以由$F(x,y)$运用概率运算的连续性求得。具体地，取$y_n,n\geqslant 1$，是一趋于∞的递增数列。则由于事件$\{X\leqslant x,Y\leqslant y_n\},n\geqslant 1$，是递增的及

$$\lim_{n\to\infty}\{X\leqslant x,Y\leqslant y_n\}=\bigcup_{n=1}^{\infty}\{X\leqslant x,Y\leqslant y_n\}=\{X\leqslant x\}$$

由连续性推得

$$\lim_{n\to\infty}P\{X\leqslant x,Y\leqslant y_n\}=P\{X\leqslant x\}$$

或等价地，

$$F_X(x)=\lim_{y\to\infty}F(x,y)$$

类似地有

$$F_Y(y)=\lim_{x\to\infty}F(x,y)$$

如果对一切 x 与 y 有

$$F(x,y)=F_X(x)F_Y(y)$$

则称随机变量 X 与 Y 是独立的。

随机变量 X 与 Y 称为联合连续的，如果存在函数 $f(x,y)$，称为联合概率密度函数，使下式对一切集 A 与 B 成立：

$$P\{X \text{ 取值于 } A, Y \text{ 取值于 } B\}=\iint_{AB} f(x,y)dydx$$

任意一组随机变量 $X_1,X_2,\cdots,X_n$ 的联合分布函数定义为

$$F(x_1,\cdots,x_n)=P(X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n)$$

再有，如果下式成立

$$F(x_1,\cdots,x_n)=F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$$

其中

$$F_{X_i}(x_i)=\lim_{\substack{x_j\to\infty\\ j\neq i}}F(x_1,\cdots,x_n)。$$

则称这 n 个随机变量是独立的。

1.3 期望

随机变量 X 的期望或均值，记作 $E(X)$，定义为

(1.3.1)
$$E(X)=\int_{-\infty}^{+\infty}xdF(x)$$
$$=\begin{cases}\int_{-\infty}^{+\infty}xf(x)dx, & \text{当 } X \text{ 是连续的}\\ \sum_x xP\{X=x\}, & \text{当 } X \text{ 是离散的}\end{cases}$$

倘若上述积分存在。

(1.3.1)式也定义了 X 的任何函数，比如 $h(X)$ 的期望。因为 $h(X)$ 本身是随机变量，从(1.3.1)得

$$E[h(X)]=\int_{-\infty}^{+\infty} x dF_h(x)$$

其中 F_h 是 $h(X)$ 的分布函数。然而，可以证明它等于 $\int_{-\infty}^{+\infty} h(x)dF(x)$。即

(1.3.2) $$E[h(X)] = \int_{-\infty}^{+\infty} h(x)dF(x)$$

上式有时被称为不自觉的统计学家的法则(因为统计学家一直被指责为用了(1.3.2)式但不明白它不是一个定义)。

随机变量 X 的方差定义为

$$\begin{aligned}\mathrm{Var}X &= E[(X-E[X])^2] \\ &= E[X^2]-E^2[X]。\end{aligned}$$

两个随机变量 X 与 Y 被称为不相关的，若它们的协方差

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E[(X-E(X))(Y-E(Y))] \\ &= E(XY)-E(X)E(Y)\end{aligned}$$

是零。由此可知，独立随机变量是不相关的，但其逆不真(读者应举出一个例子)。

期望的一个重要性质是随机变量之和的期望等于期望之和：

(1.3.3) $$E\left[\sum_{i=1}^{n} X_i\right]=\sum_{i=1}^{n} E[X_i]$$

相应的方差性质为

(1.3.4)

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right]= \sum_{i=1}^{n}\mathrm{Var}(X_i) + 2\sum_{i<j}\sum \mathrm{Cov}(X_i,X_j)$$

例 1.3(a)　匹配问题。 在一次集会上，n 个人把他们的帽子放到房间的中央混合在一起，而后每个人随机地选取一顶，我们感兴趣的是拿到自己的帽子的人数 X 的均值与方差。

为了求解，我们利用表示式

$$X=X_1+X_2+\cdots+X_n$$

其中

$$X_i=\begin{cases}1, & \text{如果第 } i \text{ 个人拿到自己的帽子} \\ 0, & \text{其它}\end{cases}$$

因为第 i 个人是等可能地选到 n 顶帽子中的任何一顶，所以得 $P\{X_i=1\}=1/n$，从而

$$E(X_i)=1/n$$
$$\text{Var}(X_i)=\frac{1}{n}\left(1-\frac{1}{n}\right)=\frac{n-1}{n^2}$$

又

$$\text{Cov}(X_i,X_j)=E(X_iX_j)-E(X_i)E(X_j)$$

现在由于

$$X_iX_j=\begin{cases}1, & \text{若第 } i \text{ 个人与第 } j \text{ 个人都拿到自己的帽子}\\ 0, & \text{其它}\end{cases}$$

得

$$\begin{aligned}E(X_iX_j)&=P\{X_i=1,X_j=1\}\\&=P\{X_i=1\}P\{X_j=1/X_i=1\}\\&=\frac{1}{n}\frac{1}{n-1}\end{aligned}$$

因此，

$$\text{Cov}(X_i,X_j)=\frac{1}{n(n-1)}-\left(\frac{1}{n}\right)^2=\frac{1}{n^2(n-1)}$$

所以，由(1.3.3)与(1.3.4)，

$$E[X]=1$$

及

$$\text{Var}(X)=\frac{n-1}{n}+2\binom{n}{2}\frac{1}{n^2(n-1)}=1$$

于是匹配数的均值与方差都是 1。(关于这些结果为何不令人惊奇的解释见例 1.5(e)。)

例 1.3(b)　某些概率恒等式。以 A_1，A_2，…，A_n 记事件，定义示性变量 I_j，$j=1$，2，…，n，为

$$I_j=\begin{cases}1, & \text{若 } A_j \text{ 发生}\\ 0, & \text{其它}\end{cases}$$

令

$$N=\sum_{j=1}^{n}I_j$$

则 N 表示 A_j，$1\leqslant j\leqslant n$，中发生的事件的个数。注意到

(1.3.5) $$(1-1)^N=\begin{cases}1, & 若 N=0\\ 0, & 若 N>0\end{cases}$$

就能得到一个有用的恒等式。由二项式定理

(1.3.6) $$\begin{aligned}(1-1)^N &= \sum_{i=0}^{N}\binom{N}{i}(-1)^i\\ &= \sum_{i=0}^{n}\binom{N}{i}(-1)^i \quad 因为\binom{m}{i}=0，当 i>m\end{aligned}$$

因此，如果我们记

$$I=\begin{cases}1, & 若 N>0\\ 0, & 若 N=0\end{cases}$$

则(1.3.5)与(1.3.6)导致

$$1-I=\sum_{i=0}^{n}\binom{N}{i}(-1)^i$$

或

(1.3.7) $$I=\sum_{i=1}^{n}\binom{N}{i}(-1)^{i+1}$$

在(1.3.7)的两边取期望得

(1.3.8) $$E[I]=E[N]-E\left[\binom{N}{2}\right]+\cdots+(-1)^{n+1}E\left[\binom{N}{n}\right]$$

然而

$$\begin{aligned}E[I]&=P\{N>0\}\\ &=P\{至少有一 A_i 发生\}\\ &=P\{\bigcup_{i=1}^{n}A_i\}\end{aligned}$$

且

$$E[N]=E\left[\sum_{j=1}^{n}I_j\right]=\sum_{j=1}^{n}P(A_j)$$

$$\begin{aligned}E\left[\binom{N}{2}\right]&=E[A_j 中两个事件同时发生的组数]\\ &=E\left[\sum\sum_{i<j}I_iI_j\right]\\ &=\sum\sum_{i<j}E[I_iI_j]\\ &=\sum\sum_{i<j}P(A_iA_j)\end{aligned}$$

一般地，同理有

$$E\left[\binom{N}{i}\right]=E[i\text{ 个事件同时发生的组数}]$$
$$=E\left[\sum\sum_{j_1<j_2<\cdots<j_i}\sum I_{j_1}I_{j_2}\cdots I_{j_i}\right]$$
$$=\sum\sum_{j_1<j_2<\cdots<j_i}\sum P(A_{j_1}A_{j_2}\cdots A_{j_i})$$

因此，(1.3.8)正是熟知的恒等式

$$P(\bigcup_{j=1}^{n}A_j)=\sum_{i=1}^{n}P(A_i)-\sum_{i<j}\sum P(A_iA_j)+\sum\sum_{i<j<k}\sum P(A_iA_jA_k)$$
$$-\cdots+(-1)^{n+1}P(A_1A_2\cdots A_n)$$

用此方法也可导出其它有用的恒等式。例如，假设我们想求事件 $A_1,\cdots,A_n$ 中恰有 r 个发生的概率的公式，那么定义

$$I_r=\begin{cases}1, & \text{若 } N=r\\ 0, & \text{其它}\end{cases}$$

且使用恒等式

$$\binom{N}{r}(1-1)^{N-r}=I_r$$

或

$$I_r=\binom{N}{r}\sum_{i=0}^{N-r}\binom{N-r}{i}\quad(-1)^i$$
$$=\sum_{i=0}^{n-r}\binom{N}{r}\binom{N-r}{i}\quad(-1)^i$$
$$=\sum_{i=0}^{n-r}\binom{N}{r+i}\binom{r+i}{r}\quad(-1)^i$$

上式两端取期望得

$$E[I_r]=\sum_{i=0}^{n-r}(-1)^i\binom{r+i}{r}E\left[\binom{N}{r+i}\right]$$

或

(1.3.9) $\quad P\{\text{事件 } A_1,\ \cdots,\ A_n \text{ 中恰有 } r \text{ 个发生}\}$

$$=\sum_{i=0}^{n-r}\ (-1)^i\binom{r+i}{r}\sum_{j_1<j_2<\cdots<j_{r+i}}\cdots\sum P(A_{j_1}A_{j_2}\cdots A_{j_{r+i}})$$

作为(1.3.9)的一个应用，假设 m 个球随机地放入 n 个盒中，每个球等可能地进入 n 个盒中的任一个，与其它球的位置相互独立。我们来计算恰有 r 个空盒的概率。以 A_i 记第 i 个盒子是空的这一事件，从(1.3.9)得

$$P\{\text{恰有 } r \text{ 个空盒}\} = \sum_{i=0}^{n-r}(-1)^i \binom{r+i}{r}\binom{n}{r+i}\left(1-\frac{r+i}{n}\right)^m$$

因和式 $\sum_{i_1<i_2<\cdots<i_{r+i}}$ 由 $\binom{n}{r+i}$ 项组成而和式中每一项都等于指定的 $r+i$ 个盒子是空的概率，故得上式。

1.4 矩母函数、特征函数及拉普拉斯变换

X 的矩母函数定义为

$$\psi(t) = E[e^{tX}]$$
$$= \int e^{tx} dF(x)$$

对 ψ 逐次求导并计算在 $t=0$ 点的值能得到 X 的各阶矩，即

$$\psi'(t) = E[Xe^{tX}]$$
$$\psi''(t) = E[X^2 e^{tX}]$$
$$\vdots$$
$$\psi^n(t) = E[X^n e^{tX}]$$

计算在 $t=0$ 点的值得

$$\psi^n(0) = E[X^n], \qquad n \geqslant 1$$

应注意我们已假定求导与积分运算可交换是合理的。通常遇到的情形都是这样的。

当矩母函数存在时，它唯一地决定分布。这是十分重要的，因为这使我们能够用矩母函数刻划随机变量的概率分布。

例 1.4(a) 设 X 与 Y 是独立的正态随机变量，各自的均值为 μ_1 与 μ_2，方差为 σ_1^2 与 σ_2^2，它们的和的矩母函数由下式给出

$$\psi_{X+Y}(t) = E[e^{t(X+Y)}]$$
$$= E[e^{tX}]E[e^{tY}] \qquad (\text{由独立性})$$
$$= \psi_X(t)\psi_Y(t)$$
$$= \exp\{(\mu_1+\mu_2)t + (\sigma_1^2+\sigma_2^2)t^2/2\}$$

其中最后的等式来自表 1.4.2，于是 $X+Y$ 的矩母函数是均值为 $\mu_1+\mu_2$，方差为 $\sigma_1^2+\sigma_2^2$ 的正态随机变量的矩母函数。由唯一性，这就是 $X+Y$ 的分布。

表 1.4.1

离散概率分布	概率质量函数 $p(x)$	矩母函数 $\phi(t)$	均值	方差
二项分布参数 $n,p,0\leqslant p\leqslant 1$	$\binom{n}{x}p^x(1-p)^{n-x}$ $x=0,1,\cdots,n$	$[pe^t+(1-p)]^n$	np	$np(1-p)$
泊松分布 参数 $\lambda>0$	$e^{-\lambda}\frac{\lambda^x}{x!}$ $x=0,1,2,\cdots$	$\exp\{\lambda(e^t-1)\}$	λ	λ
几何分布 参数 $0\leqslant p\leqslant 1$	$p(1-p)^{x-1}$ $x=1,2,\cdots$	$\frac{pe^t}{1-(1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
负二项分布 参数 r,p	$\binom{x-1}{r-1}p^r(1-p)^{x-r}$ $x=r,r+1,\cdots$	$\left[\frac{pe^t}{1-(1-p)e^t}\right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

表 1.4.2

连续概率分布	概率密度函数 $f(x)$	矩母函数 $\phi(t)$	均值	方差
(a,b)上的均匀分布	$\frac{1}{b-a},\quad a<x<b$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布 参数 $\lambda>0$	$\lambda e^{-\lambda x},\quad x\geqslant 0$	$\frac{\lambda}{\lambda-t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Γ-分布参数 $(n,\lambda),\lambda>0$	$\frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}$ $x\geqslant 0$	$\left(\frac{\lambda}{\lambda-t}\right)^n$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
正态分布 参数(μ,σ^2)	$\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$ $-\infty<x<\infty$	$\exp\left\{\mu t+\frac{\sigma^2t^2}{2}\right\}$	μ	σ^2
B-分布 参数 a,b, $a>0,b>0$	$cx^{a-1}(1-x)^{b-1}$ $0<x<1$ $c=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$		$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2/(a+b+1)}$

由于一个随机变量 X 的矩母函数不一定存在，故理论上更方便的是定义 X 的特征函数：

$$\phi(t)=E[e^{itX}], \qquad -\infty<t<\infty$$

其中 $i=\sqrt{-1}$，能够证明 ϕ 总是存在的，且像矩母函数一样唯一地决定 X 的分布。

我们也可以定义随机变量 X_1，…，X_n 的联合矩母函数为

$$\psi(t_1,\cdots,t_n)=E[\exp\{\sum_{j=1}^{n}t_jX_j\}]$$

或联合特征函数为

$$\phi(t_1,\cdots,t_n)=E[\exp\{i\sum_{j=1}^{n}t_jX_j\}]$$

可以证明联合矩母函数(当它存在时)或联合特征函数唯一地决定联合分布。

在处理只取非负值的随机变量时，使用拉普拉斯变换比起特征函数来有时更为方便。分布 F 的拉普拉斯变换定义为

$$\widetilde{F}(s)=\int_0^\infty e^{-sx}dF(x)$$

对复数 $s=a+ib$，其中 $a\geqslant 0$，这个积分存在。正如特征函数一样，拉普拉斯变换唯一决定分布。

我们也可以对任意函数定义拉普拉斯变换如下：函数 g 的拉普拉斯变换，记作 $\widetilde{g}$，定义为

$$\widetilde{g}(s)=\int_0^\infty e^{-sx}dg(x)$$

倘若此积分存在。能证明 $\widetilde{g}$ 可确定 g，只相差一个常数。

1.5 条件期望

如果 X 与 Y 是离散随机变量，对一切使 $P\{Y=y\}>0$ 的 y，定义给定 $Y=y$ 时，X 的条件概率质量函数为

$$P\{X=x\mid Y=y\}=\frac{P\{X=x,Y=y\}}{P\{Y=y\}}$$

给定 $Y=y$ 时，X 的条件分布定义为

$$F(x|y)=P\{X\leqslant x|Y=y\}$$

而给定 $Y=y$ 时，X 的条件期望为

$$E[X|Y=y]=\int x dF(x|y)=\sum_x xP\{X=x|Y=y\}$$

如果 X 与 Y 有联合概率密度函数 $f(x,y)$，则对一切使 $f_Y(y)\geqslant 0$的 y，给定 $Y=y$ 时，X 的条件概率密度函数定义为

$$f(x|y)=\frac{f(x,y)}{f_Y(y)}$$

给定 $Y=y$ 时，X 的条件概率分布函数为

$$F(x|y)=P\{X\leqslant x|Y=y\}=\int_{-\infty}^{x}f(x|y)dx$$

而给定 $Y=y$ 时，X 的条件期望此时定义为

$$E[X|Y=y]=\int x dF(x|y)=\int xf(x|y)dx$$

所以除了概率现在都是关于事件 $Y=y$ 的条件概率外，一切定义均与无条件的情形一样。

我们以 $E[X|Y]$ 表示随机变量 Y 的函数，它在 $Y=y$ 时，取值为 $E[X|Y=y]$。条件期望的一个极其有用的性质是对一切随机变量 X 与 Y，当期望存在时，有

(1.5.1) $$E[X]=E[E[X|Y]]=\int E[X|Y=y]dF_Y(y)$$

如果 Y 是一个离散变量，则（1.5.1）式为

$$E[X]=\sum_y E[X|Y=y]P\{Y=y\}$$

而如果 Y 是连续的，具有密度 $f(y)$时，则(1.5.1)式为

$$E[X]=\int_{-\infty}^{+\infty}E[X|Y=y]f(y)dy$$

现在我们就 X 与 Y 都是离散随机变量的情形给出(1.5.1)式的证明。

(1.5.1)在 X 与 Y 是离散时的证明 要证明

$$E[X]=\sum_y E[X|Y=y]P\{Y=y\}$$

把上式的右边写为

$$\begin{aligned}
&\sum_y E[X|Y=y]P\{Y=y\} \\
&= \sum_y \sum_x xP\{X=x|Y=y\}P\{Y=y\} \\
&= \sum_y \sum_x xP\{X=x,Y=y\} \\
&= \sum_x x\sum_y P\{X=x,Y=y\} \\
&= \sum_x xP\{X=x\} \\
&= E[X]
\end{aligned}$$

结论得证。

从(1.5.1)式我们看到，$E[X]$是给定$Y=y$时，X的条件期望的一个加权平均值，每一项$E[X|Y=y]$所加的权是作为条件的事件的概率。

例 1.5(a)　随机个随机变量之和。设X_1，X_2，…，为一列独立同分布的随机变量；设N为一非负整值随机变量，且与序列X_1，X_2，…，独立。我们首先在对N取条件的情况下来计算$Y=\sum_{i=1}^{N}X_i$的矩母函数。即

$$\begin{aligned}
&E[\exp\{t\sum_{i=1}^{N}X_i\}|N=n] \\
&\quad = E[\exp\{t\sum_{i=1}^{n}X_i\}|N=n] \\
&\quad = E[\exp\{t\sum_{i=1}^{n}X_i\}] \quad \text{（由独立性）} \\
&\quad = (\psi_X(t))^n
\end{aligned}$$

其中$\psi_X(t)=E[e^{tX}]$是X的矩母函数。因此，

$$E[\exp\{t\sum_{i=1}^{N}X_i\}|N]=(\psi_X(t))^N$$

从而

$$\psi_Y(t)=E[\exp\{t\sum_{i=1}^{N}X_i\}]=E[(\psi_X(t))^N]$$

为了计算　$Y=\sum_{i=1}^{N}X_i$的均值与方差，我们对$\psi_Y(t)$求导：

$$\psi_Y'(t) = E[N(\psi_X(t))^{N-1}\psi_X'(t)]$$

$$\psi_Y''(t) = E[N(N-1)(\psi_X(t))^{N-2}(\psi_X'(t))^2 + N(\psi_X(t))^{N-1}\psi_X''(t)]$$

计算在 $t=0$ 点的值，得

$$E(Y) = E[NE(X)] = E[N]E[X]$$

及

$$E[Y^2] = E[N(N-1)E^2[X] + NE[X^2]] = E[N]\mathrm{Var}(X) + E[N^2]E^2\{X]$$

因此

$$\mathrm{Var}(Y) = E[Y^2] - E^2[Y] = E[N]\mathrm{Var}(X) + E^2[X]\mathrm{Var}(N)$$

例 1.5(b) 一名矿工陷进一个有三扇门的矿井中。第一扇门通到一个隧道，走两小时后他可到达安全区。第二扇门通到又一个隧道，走三小时会使他回到这矿井中。第三扇门通到再一个隧道，走五小时后，也会使他回到矿井中。假定这矿工总是等可能地在三扇门中选择一扇，让我们计算矿工到达安全区的时间 X 的矩母函数。

设 Y 表示最初选择的门。那么

(1.5.2) $$E[e^{tX}] = \frac{1}{3}\{E[e^{tX}|Y=1] + E[e^{tX}|Y=2] + E[e^{tX}|Y=3]\}$$

现在给定 $Y=1$，得 $X=2$，因而

$$E[e^{tX}|Y=1] = e^{2t}$$

又给定 $Y=2$，得 $X=3+X'$，其中 X' 是回到矿井后再找到安全区所附加的时间。但一旦矿工回到他的陷井，问题又与以前完全相同，于是 X' 与 X 有相同的分布。所以，

$$E[e^{tX}|Y=2] = E[e^{t(3+X')}] = e^{3t}E[e^{tX}]$$

类似地有，

$$E[e^{tX}|Y=3] = e^{5t}E[e^{tX}]$$

代回到(1.5.2)得

$$E[e^{tX}] = \frac{1}{3}\{e^{2t} + e^{3t}E[e^{tX}] + e^{5t}E[e^{tX}]\}$$

或

$$E[e^{tX}] = \frac{e^{2t}}{3 - e^{3t} - e^{5t}}$$

先对一个适当的随机变量取条件，不仅使我们能求得期望，也可以用这种方法计算概率。为明瞭这点，以 E 记一个任意的事件且定义示性随机变量 X 为

$$X=\begin{cases}1, & \text{如果 } E \text{ 发生}\\ 0, & \text{如果 } E \text{ 不发生}\end{cases}$$

从 X 的定义得到

$$E[X]=P(E)$$

$$E[X|Y=y]=P(E|Y=y), \quad \text{对任意的随机变量 } Y$$

所以，从(1.5.1)式我们得到

$$P(E) = \int P(E|Y = y)dF_Y(y)$$

例 1.5(c)　假定 X 与 Y 是独立随机变量，各自的分布为 F 与 G，那么 $X+Y$ 的分布，记为 $F*G$，称为 F 与 G 的卷积，由下式给出

$$\begin{aligned}(F*G)(a) &= P\{X + Y \leqslant a\}\\ &= \int_{-\infty}^{+\infty} P\{X + Y \leqslant a|Y = y\}dG(y)\\ &= \int_{-\infty}^{+\infty} P\{X + y \leqslant a|Y = y\}dG(y)\\ &= \int_{-\infty}^{+\infty} F(a - y)dG(y)\end{aligned}$$

我们以 F_2 记 $F*F$，一般地 $F*F_{n-1}=F_n$。于是 F_n，即 F 自身的 n 重卷积，是 n 个独立的、每一个分布都为 F 的随机变量之和的分布。

例 1.5(d)　**选票问题**。在一次选举中，候选人 A 得到 n 张选票而候选人 B 得到 m 张选票，其中 $n>m$，假定选票的一切排列次序是等可能的，证明：在计票过程中，A 的票数始终领先的概率为 $(n-m)/(n+m)$。

解　以 $P_{n,m}$ 记所要求的概率。以得到最后那张选票的候选人为条件，我们有

$$\begin{aligned}P_{n,m} &= P\{A \text{ 始终领先 } |A \text{ 得到最后一票}\} \frac{n}{n + m}\\ &\quad + P\{A \text{ 始终领先 } |B \text{ 得到最后一票}\} \frac{m}{n + m}\end{aligned}$$

现在容易看出，在 A 得最后一票的条件下，A 始终领先的概率与 A 得到 $(n-1)$ 票而 B 得到 m 张票的情形要算的概率是一样的。在 B 得到最后一票的条件下，可得类似的结果。从上面的讨论我们看到

(1.5.3) $$P_{n,m}=\frac{n}{n+m}P_{n-1,m}+\frac{m}{m+n}P_{n,m-1}$$

现在用归纳法，对 $n+m$ 进行归纳，能够证明

$$P_{n,m}=\frac{n-m}{n+m}$$

当 $n+m=1$ 时显然结论为真，即 $P_{1,0}=1$。假定 $n+m=k$ 时上式成立，则当 $n+m=k+1$时，由(1.5.3)与归纳假设有

$$\begin{aligned}P_{n,m}&=\frac{n}{n+m}\frac{n-1-m}{n-1+m}+\frac{m}{m+n}\frac{n-m+1}{n-m-1}\\&=\frac{n-m}{n+m}\end{aligned}$$

结论得证。

选票问题有一些有趣的应用。例如，考虑连续掷一枚硬币，其正面出现的概率总是 p。让我们确定开始掷币之后正面出现的总次数与反面出现的总次数首次相等的时刻的概率分布。首次相等的时刻为 $2n$ 的概率能先对前 $2n$ 次试验中正面出现的次数取条件而得到，这就有

$$\begin{aligned}&P\{\text{首次相等的时刻}=2n\}\\&\quad=P\{\text{首次相等的时刻}=2n\mid\text{在前 }2n\text{ 次中正面出现 }n\text{ 次}\}\binom{2n}{n}p^n(1-p)^n\end{aligned}$$

在前 $2n$ 次抛掷中出现 n 次正面的条件下，易知 n 个正面与 n 个反面的一切可能的排列是等可能的。于是上面的条件概率等于在一次选举中两候选人各得 n 张票且其中一个在计票过程中总是领先直至最后一票（这一票才使票数相等）的概率。对无论是谁得到最后一票取条件，这正是选票问题中在 $m=n-1$ 时的概率。因此

$$\begin{aligned}P\{\text{首次相等的次数}=2n\}&=P_{n,n-1}\binom{2n}{n}p^n(1-p)^n\\&=\frac{\binom{2n}{n}p^n(1-p)^n}{2n-1}\end{aligned}$$

例 1.5(e)　匹配问题(续)。让我们再考虑例 1.3(a)，即 n 个人把他们的帽子混在一起，而后每人随机地选一顶，我们要计算恰有 k 个匹配的概率。

先以 E 记全不匹配这一事件及 $P_n=P(E)$，使得明显地表明它依赖于 n。对第一个人选到或未选到自己的帽子——记这两个事件为 M 和 M^c——取条件我们得到

$$P_n = P(E) = P(E|M)P(M) + P(E|M^c)P(M^c)$$

显然，$P(E|M)=0$，从而

(1.5.4) $$P_n=P(E|M^c)\ \frac{n-1}{n}$$

现在 $P(E|M^c)$ 是 $n-1$ 个人从 n-1 顶帽子中各取一顶都不匹配的概率，其中有一个人的帽子不在这 $n-1$ 顶帽子中。此事件以两种互不相容的方式发生。一种是都不匹配且多余的那个人（即其帽子已给第一个人取走的那个人）未能选中多余的帽子（即第一个选取的人的帽子），另一种是都不匹配但多余的人选取到了多余的帽子。前一种的概率是 P_{n-1}，因为可以将多余的帽子看作为多余的人的。由于第二种的概率是 $\left[\frac{1}{n-1}\right]P_{n-2}$，我们有

$$P(E|M^c) = P_{n-1} + \frac{1}{n-1}P_{n-2}$$

于是，从(1.5.4)式得

$$P_n = \frac{n-1}{n}P_{n-1} + \frac{1}{n}P_{n-2}$$

或等价地

(1.5.5) $$P_n-P_{n-1}=-\frac{1}{n}(P_{n-1}-P_{n-2})$$

然而，显然有

$$P_1=0,\qquad P_2=\frac{1}{2}$$

于是，从(1.5.5)式得

$$P_3 - P_2 = -\frac{P_2 - P_1}{3} = -\frac{1}{3!} \quad 或 \quad P_3 = \frac{1}{2!} - \frac{1}{3!}$$

$$P_4 - P_3 = -\frac{P_3 - P_2}{4} = \frac{1}{4!} \quad 或 \quad P_4 = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}$$

而一般地，有

$$P_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{(-1)^n}{n!}$$

为了得到恰有 k 个匹配的概率，我们考虑任意固定的 k 个人，只有他们选中自己的帽子的概率是

$$\frac{1}{n}\cdot\frac{1}{n-1}\cdots\frac{1}{n-(k-1)}P_{n-k}=\frac{(n-k)!}{n!}P_{n-k}$$

其中 P_{n-k}是其余 $n-k$ 个人从他们自己的那些帽子中选取但全不匹配的概率。因 k 个人的选择法有$\binom{n}{k}$种，所以恰有 k 个匹配的概率是

$$\binom{n}{k}\frac{(n-k)}{n!}P_{n-k}=\frac{1}{k!}\left[\frac{1}{2!}-\frac{1}{3!}+\cdots+\frac{(-1)^{n-k}}{(n-k)!}\right]$$

当 n 充分大时上式近似地等于 $e^{-1}/k!$。

所以当 n 充分大时，匹配数近似地具有均值为 1 的泊松分布。为了更好地理解这个结果，我们回想一下，均值为 λ 的泊松分布正是 n 次独立试验中成功次数的极限公布，其中每次成功的概率为 p_n，当 $n\to\infty$时，$np_n\to\lambda$。现在如果我们设

$$X_i=\begin{cases}1, & \text{若第 } i \text{ 个人选到自己的帽子}\\ 0, & \text{其它}\end{cases}$$

则匹配数 $\sum_{i=1}^{n}X_i$ 可以认为是 n 次试验中成功的次数，各次成功的概率为$\frac{1}{n}$。然而上述结果不能直接运用，因为这些试验并不独立。但这里相依性是相当弱的，因为，例如

$$P\{X_i=1\}=1/n$$
$$P\{X_i=1|Y_j=1\}=1/(n-1),\qquad j\neq i$$

因此我们肯定会希望在这种弱相依关系下，泊松极限仍然成立。本例的结果表明确实如此。

例 1.5(f)　装填问题。设有 n 个点排列在直线上，且设随机地选取一对相邻的点，即选取（i，$i+1$）的概率为 1/（$n-1$），$i=1,2,\cdots,n-1$。我们连续地随机取这样的点对，但如果取到的点对中有前面已取到过的点，这一对就去掉不算。这样一直取到只留下孤立点为止。我们关心的是孤立点个数的平均数。

例如，若 $n=8$，而依次出现的随机对为（2，3），（7，8），（3，4）及（4，5），那么有两个孤立点（（3，4）这对去掉），如图 1.5.1 所示。



图 1.5.1

若我们令

$$I_{i,n}=\begin{cases}1, & 若 i 是孤立点\\ 0, & 其它\end{cases}$$

则 $\sum_{i=1}^{n} I_{i,n}$ 表示孤立点的个数。因此，

$$E[孤立点的个数] = \sum_{i=1}^{n} E[I_{i,n}] = \sum_{i=1}^{n} P_{i,n}$$

其中 $P_{i,n}$ 为有 n 个点时，i 是孤立点的概率。令

$$P_n \equiv P_{n,n} = P_{1,n}$$

P_n 是端点 n（或 1）成孤立点的概率。为了导出 $P_{i,n}$ 的表达式，可以认为这 n 个点由两个独立的片断组成，即

$$1, 2, \cdots, i \quad 与 \quad i, i+1, \cdots, n。$$

因为点 i 未取到当且仅当第一段的右端点和第二段的左端点都未取到，所以得

(1.5.6) $$P_{i,n} = P_i P_{n-i+1}$$

因此如果我们能计算端点未取到的概率，则 $P_{i,n}$ 可确定。为导出 P_n 的表达式，以初始对（比如说 $(i, i+1)$）为条件，且注意这个点对将这些点断成两个独立的片断——$1, 2, \cdots, i-1$ 及 $i+2, \cdots, n$。于是，如果初始点对是 $(i, i+1)$，那么端点 n 是孤立点，如果 $(n-i-1)$ 个点的集合的端点是孤立点。因此有

$$P_n = \sum_{i=1}^{n-1} \frac{P_{n-i-1}}{n-1} = \frac{P_1 + \cdots + P_{n-2}}{n-1}$$

或

$$(n-1)P_n = P_1 + \cdots + P_{n-2}$$

用 $n-1$ 代替 n 得

$$(n-2)P_{n-1} = P_1 + \cdots + P_{n-3}$$

将两式相减得

$$(n-1)P_n - (n-2)P_{n-1} = P_{n-2}$$

或

$$P_n - P_{n-1} = -\frac{1}{n-1}(P_{n-1} - P_{n-2})$$

由 $P_1=1$ 及 $P_2=0$，得

$$P_3 - P_2 = -\frac{P_2 - P_1}{2} = \frac{1}{2} \quad \text{或} \quad P_3 = \frac{1}{2!}$$

$$P_4 - P_3 = -\frac{1}{3}(P_3 - P_2) = -\frac{1}{3!} \quad \text{或} \quad P_4 = \frac{1}{2!} - \frac{1}{3!}$$

一般地，

$$P_n = \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{n-1}}{(n-1)!} = \sum_{j=0}^{n-1} \frac{(-1)^j}{j!}, \quad n \geqslant 2$$

于是，从(1.5.6)得

$$P_{i,n} = \begin{cases} \displaystyle\sum_{j=0}^{n-1} \frac{(-1)^j}{j!}, & i = 1, n \\ 0, & i = 2, n-1 \\ \displaystyle\sum_{j=0}^{i-1} \frac{(-1)^j}{j!} \sum_{j=0}^{n-i} \frac{(-1)^j}{j!}, & 2 < i < n-1 \end{cases}$$

对大的 i 与 $n-i$，由上面可见 $P_{i,n} \approx e^{-2}$，且事实上能从上式证明 $\sum_{i=1}^{n} P_{i,n}$（孤立点的期望数）渐近地为

$$\sum_{i=1}^{n} P_{i,n} \approx (n+2)e^{-2}, \quad \text{对充分大的 } n$$

例 1.5(g)　可靠性一例。考虑一个有 n 个部件的系统，它受到随机地发生的冲击，假定每次冲击的大小(用一个数值表示)服从分布 G，与其它的冲击值相互独立。若发生一次值为 x 的冲击，则冲击到达时，正在工作的各部件相互独立地以概率 x 在瞬时间失效 * 。我们关心的是直到全体部件都失效所必需的冲击次数 N 的分布。

为计算 $P\{N>k\}$，以 E_i，$i=1, 2, \cdots, n$，记部件 i 在前 k 次冲击后仍幸存完好这一事件，则

$$\begin{aligned} P\{N>k\} &= P\Big(\bigcup_{i=1}^{n} E_i\Big) \\ &= \sum_{i=1}^{n} P(E_i) - \sum_{i<l} P(E_i E_l) + \cdots + (-1)^{n+1} P(E_1 E_2 \cdots E_n) \end{aligned}$$

为了计算此概率，以 p_j 记任意一次冲击后，指定的 j 个部件全部完好的概率，以冲击值为条件得

---

* 由概率的意义，这要求冲击的值在区间(0,1)之中。——译者注

$$p_j = \int P\{\text{指定的 } j \text{ 个幸存完好} \mid \text{冲击值是 } x\} dG(x)$$

$$= \int (1-x)^j dG(x)$$

因为

$$P(E_i) = p_1^k$$

$$P(E_i E_l) = p_2^k, \cdots, P(E_1 E_2 \cdots E_n) = p_n^k$$

可见

$$P\{N > k\} = np_1^k - \binom{n}{2} p_2^k + \binom{n}{3} p_3^k + \cdots + (-1)^{n+1} p_n^k$$

由此可计算 N 的均值如下：

$$\begin{aligned} E[N] &= \sum_{k=0}^{\infty} P\{N > k\} \\ &= \sum_{k=0}^{\infty} \sum_{i=1}^{n} \binom{n}{i} (-1)^{i+1} p_i^k \\ &= \sum_{i=1}^{n} \binom{n}{i} (-1)^{i+1} \sum_{k=0}^{\infty} p_i^k \\ &= \sum_{i=1}^{n} \binom{n}{i} \frac{(-1)^{i+1}}{1-p_i} \end{aligned}$$

读者应注意到我们已使用了恒等式 $E[N] = \sum_{k=0}^{\infty} P\{N > k\}$，它对一切非负整值随机变量 N 成立（见习题 1）。

1.6 指数分布、无记忆性及失效率函数

一个连续随机变量 X 的概率密度函数若为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

或等价地其分布为

$$F(x) = \int_{-\infty}^{x} f(y) dy = \begin{cases} 1 - e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

则称 X 具有参数为 $\lambda(\lambda>0)$ 的指数分布。

指数分布的矩母函数为

$$E[e^{tX}]=\int_0^{\infty}e^{tx}\lambda e^{-\lambda x}dx=\frac{\lambda}{\lambda-t} \tag{1.6.1}$$

现在能对(1.6.1)求导而得到 X 的各阶矩，我们留给读者去验证

$$E[X]=1/\lambda,\quad \mathrm{Var}(X)=1/\lambda^2$$

指数随机变量之所以有用是因为它们具有无记忆性，这里一个随机变量 X 称为无记忆的，若

$$P\{X>s+t|X>t\}=P\{X>s\},\quad s,t\geqslant 0 \tag{1.6.2}$$

如果我们把 X 看作某仪器的寿命，那么 (1.6.2) 的意思是在仪器已工作了 t 小时的条件下，它至少工作 $s+t$ 小时的概率与它原来至少工作 s 小时的概率是相同的。换句话说，如果仪器在时刻 t 是完好的，则它的剩余寿命的分布就是原来的寿命的分布。条件(1.6.2)等价于

$$\overline{F}(s+t)=\overline{F}(s)\overline{F}(t)$$

因为当 F 是指数分布时，上式成立，可见这样的随机变量是无记忆的。

例 1.6(a) 考虑一个有两名营业员的邮局。假设当 A 进去时，他发现一名营业员正在给 B 办事而另一名则正在为 C 服务。还假设已告诉 A，一旦 B 或 C 离开就为他服务。如果一个办事员为一个顾客所花的时间服从均值是 $1/\lambda$ 的指数分布。三个顾客中 A 最后离开邮局的概率是多少?

由如下推理可得解答：考虑 A 发现一个营业员有空的时刻，此时 B 与 C 中有一个刚好离开而另一个仍在接受服务. 然而由指数分布的无记忆性，这另一个人在邮局再要花费的时间也服从指数分布，其均值仍为 $1/\lambda$，即仿佛他才开始服务。因此由对称性，他在 A 之前结束服务的概率为 $\frac{1}{2}$。

例 1.6(b) 设 $X_1, X_2, \cdots$ 为独立同分布的连续随机变量，具有分布 F。如果 $X_n>\max(X_1,\cdots,X_{n-1}), n>0$ (其中 $X_0=-\infty$)，则说在时刻 n 产生一个记录，其值为 X_n，也就是每次产生一个记录，就达到一个新的高度。设 τ_i 表示第 i 个与第 $i+1$ 个记录之间的时间，其分布是什么?

作为计算 τ_i 的分布的第一步，我们注意到序列 X_1，X_2，$\cdots$ 与 $F(X_1)$，

$F(X_2)$,…的记录时刻是一样的，且因 $F(X)$ 服从(0,1)上的均匀分布(见习题2)，所以 τ_i 的分布不依赖于具体的分布 F(只需它是连续的)。因此，我们假定 F 是参数 $\lambda=1$ 的指数分布。

为计算 τ_i 的分布，我们将以第 i 个记录值 R_i 为条件。现在 $R_1=X_1$ 是参数为 1 的指数变量。在 R_2 比 R_1 大的条件下，R_2 有参数为 1 的指数分布。由指数变量的无记忆性可得，R_2 的分布等同于 R_1 与一个参数为 1 的独立指数变量之和的分布。因此 R_2 的分布等同于两个独立的参数为 1 的指数变量之和的分布。由同样的推理可知 R_i 的分布等同于 i 个独立的参数为 1 的指数变量之和的分布。但众所周知(见习题 17)，这样的随机变量具有参数为 $(i,1)$ 的 Γ-分布，即 R_i 的密度为

$$f_{R_i}(t)=\frac{e^{-t}t^{i-1}}{(i-1)!},\qquad t\geqslant 0$$

因此，以 R_i 为条件得

$$\begin{aligned}P\{\tau_i>k\}&=\int_0^\infty P\{\tau_i>k\,|\,R_i=t\}\frac{e^{-t}t^{i-1}}{(i-1)!}dt\\&=\int_0^\infty(1-e^{-t})^k e^{-t}\frac{t^{i-1}}{(i-1)!}dt,\qquad i\geqslant 1\end{aligned}$$

因为当第 i 个记录值为 t 时，接着的 k 个值都不是记录，如果它们都小于 t，由此推得最后的等式。

不但指数分布是无记忆性的，而且它是唯一具有这个性质的分布。为证此，假设 X 是无记忆的*)，且记 $\overline{F}(x)=P\{X>x\}$，则

$$\overline{F}(s+t)=\overline{F}(s)\cdot\overline{F}(t)$$

即 $\overline{F}(x)$ 满足函数方程

$$g(s+t)=g(s)g(t)$$

然而上述方程满足某种适当的条件(如单调性，右或左连续性，甚至可测性)的解只能有如下形式

$$g(x)=e^{-\lambda x}$$

λ 是某个适当的数值。(假定 g 右连续时，一个简单的证明如下：因 $g(s+t)=g(s)g(t)$，得 $g(2/n)=g(1/n+1/n)=g^2(1/n)$。反复推知 $g(m/n)=g^m(1/n)$。又 $g(1)=g(1/n+1/n+\cdots+1/n)$

*) 本节只讨论非负随机变量。—— 译者注

$= g^n(1/n)$。因此，$g(m/n) = [g(1)]^{\frac{m}{n}}$。既然 g 右连续，由此蕴含 $g(x) = [g(1)]^x$。因 $g(1) = g^2(1/2) \geqslant 0$，得到 $g(x) = e^{-\lambda x}$，其中 $\lambda = -\log(g(1))$。）因为分布函数总是右连续的，所以必有

$$\overline{F}(x) = e^{-\lambda x}$$

指数变量的无记忆性可由指数分布的失效率函数（也称风险率函数）进一步予以阐明。

考虑一个连续随机变量 X，它有分布函数 F 与密度函数 f。失效（或风险）率函数 $\lambda(t)$ 定义为

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)} \tag{1.6.3}$$

为了阐明 $\lambda(t)$ 的意义，把 X 设想为某种元件的寿命，且假定 X 已经存活 t 小时，我们要求再过时间 dt 它失效的概率，即考虑 $P\{X \in (t,t+dt) \mid X > t\}$。现在

$$\begin{aligned}
P\{X \in (t,t+dt) \mid X > t\} &= \frac{P\{X \in (t,t+dt), X > t\}}{P\{X > t\}} \\
&= \frac{P\{X \in (t,t+dt)\}}{P\{X > t\}} \\
&\approx \frac{f(t)dt}{\overline{F}(t)} \\
&= \lambda(t)dt
\end{aligned}$$

可见 $\lambda(t)$ 表示一个 t 岁的元件将失效的概率强度。

现在假设寿命分布是指数分布，那么由无记忆性，一个 t 岁的元件的剩余寿命的分布与一个新元件的寿命分布相同。因此 $\lambda(t)$ 应当是常数。这从下式可得证

$$\lambda(t) = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

于是指数分布的失效率函数是常数。参数 λ 常称为分布的速率。（注意速率是均值的倒数，反之亦然。）

失效率函数 $\lambda(t)$ 唯一决定分布 F。为证明此事，我们注意到

$$\lambda(t) = \frac{-\frac{d}{dt}\overline{F}(t)}{\overline{F}(t)}$$

积分得

$$\log\overline{F}(x) = -\int_0^t \lambda(t)dt + k$$

或

$$\overline{F}(t) = c \cdot \exp\left\{-\int_0^t \lambda(t)dt\right\}$$

令 $t=0$ 得 $c=1$,因而

$$\overline{F}(t) = \exp\left\{-\int_0^t \lambda(t)dt\right\}$$

1.7 极限定理

概率论中的一些最重要的结果是以极限定理的形式出现的。两条最重要的是:

强大数定律

如果 $X_1,X_2,\cdots$独立同分布,具有均值 μ,则

$$P\{\lim_{n\to\infty}(X_1+\cdots+X_n)/n=\mu\}=1$$

中心极限定理

若 $X_1,X_2,\cdots$独立同分布,具有均值 μ 与方差 σ^2,则

$$\lim_{n\to\infty}P\left\{\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\leqslant a\right\}=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

如果我们令 $S_n=\sum_{i=1}^{n}X_i$,其中 $X_1,X_2,\cdots$独立同分布,则强大数定律说 S_n/n 以概率 1 收敛于 $E[X_i]$;而中心极限定理说当 $n\to\infty$时,S_n 有渐近正态分布。

1.8 随机过程

一个随机过程$X=\{X(t),t\in T\}$是一族随机变量,即对指标集 T 中的每个 t,$X(t)$是一个随机变量。我们常常把 t 解释成时间,且

称 $X(t)$是过程在时刻 t 的状态。如果指标集 T 是一个可数集,则称$\underline{X}$ 为一个离散时间的随机过程,而如果 T 是一个连续统,则称它为连续时间过程。

$\underline{X}$ 的任一现实称为一条样本路径。例如,事件随时间随机地发生,而 $X(t)$表示$[0,t]$中发生的事件个数,则图 1.8.1 给出$\underline{X}$ 的一条样本路径,它对应于在时刻 1 发生初始事件,在时刻 3 发生第二个事件,而在时刻 4 发生第三个,此外没有事件发生。

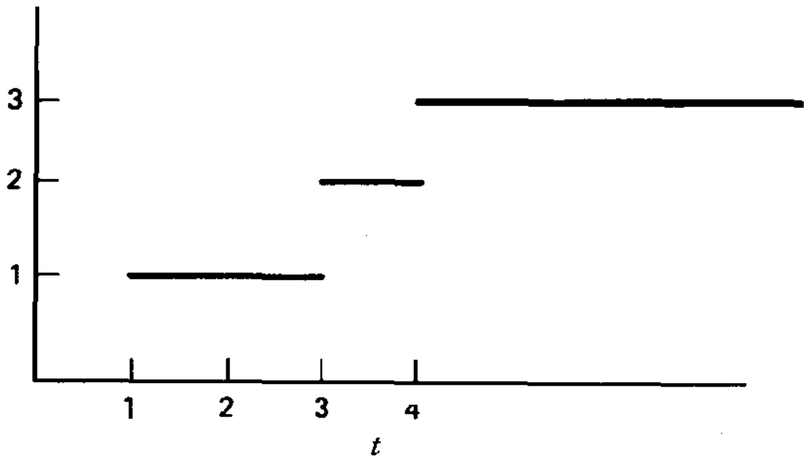


图 1.8.1 $X(t)=[0,t]$中事件的个数的一条样本路径 BZ]

连续时间随机过程$\{X(t),t\in T\}$称为有独立增量,若对一切 $t_0<t_1<\cdots<t_n$,随机变量

$$X(t_1)-X(t_0),X(t_2)-X(t_1),\cdots,X(t_n)-X(t_{n-1})$$

相互独立。称它有平稳增量,如果 $X(t+s)-X(t)$ 对一切 t 有相同的分布。也就是过程在不相重迭的区间上的增量独立时,过程有独立增量;而过程在任意两点间的增量的分布仅依赖于该两点间的距离时,过程有平稳增量。

习 题

1. 设 N 为一非负整值随机变量。证明:

$$E[N]=\sum_{k=1}^{\infty}P\{N\geqslant k\}=\sum_{k=0}^{\infty}P\{N>k\}$$

一般地，若 X 是非负的，具有分布 F，则

$$E[X]=\int_0^\infty \overline{F}(x)dx$$

及

$$E[X^n]=\int_0^\infty nx^{n-1}\overline{F}(x)dx$$

2. 设 X 是一连续随机变量，具有分布 F，证明：

(a) $F(X)$服从(0,1)上的均匀分布；

(b) 如果 U 是(0,1)上均匀分布的变量，则 $F^{-1}(U)$有分布 F，其中 $F^{-1}(x)$是满足$F(y)=x$ 的 y 值。

3. 设 X_n 为具有参数(n,p_n)，$n\geqslant 1$，的二项分布随机变量。如果当$n\to\infty$时，$np_n\to\lambda$，证明：

$$P\{X_n=i\}\to e^{-\lambda}\lambda^i/i\,!\,,\quad n\to\infty$$

4. 计算具有参数 n 及 p 的二项分布随机变量的均值与方差。

5. 假定进行 n 次独立试验，每次试验的结果为 $1,2,\cdots,r$ 中的一个，它们出现的概率分别为 $p_1,p_2,\cdots,p_r$，$\sum_{i=1}^{r}p_i=1$。以 N_i 记试验结果为 i 的次数。

(a) 计算 $N_1,\cdots,N_r$ 的联合分布。它称为多项分布。

(b) 计算 $\mathrm{Cov}(N_i,N_j)$。

(c) 计算未出现的结果的个数的均值与方差。

6. 设 $X_1,X_2,\cdots$ 是独立同分布的连续随机变量。如果 $X_n>\max(X_1,\cdots,X_{n-1})$，$n>0$，$X_0\equiv-\infty$，我们说在时刻 n 产生一个记录且值为 X_n。

(a) 以 N_n 记到(包括)时刻 n 为止产生的记录的个数，计算 $E[N_n]$及 $\mathrm{Var}(N_n)$。

(b) 令 $T=\min\{n:n>1$ 且在时刻 n 产生一个记录$\}$，计算 $P\{T>n\}$，并证明 $P(T<\infty)=1$ 及 $E[T]=\infty$。

(c) 以 T_y 记第一个大于 y 的记录值的时刻，即

$$T_y=\min\{n:X_n>y\}$$

证明：T_y 与 X_{T_y}独立，即值首次大于 y 的时刻与该值独立。(如果你反复思考最后这个命题，它似乎还是比较直观的。)

7. 以 X 记从装有 n 个白球及 m 个黑球的罐内任取 k 个球中白球的个数。计算 $E[X]$与 $\mathrm{Var}(X)$。

8. 设 X_1 与 X_2 是独立的泊松分布随机变量，均值各为 λ_1 及 λ_2。

(a) 求 X_1+X_2 的分布。

(b) 计算在 $X_1+X_2=n$ 的条件下 X_1 的条件分布。

9. 掷一枚硬币，其正面出现的概率为 p，连续地投掷直到一连出现了 r 个正面为止，计算投掷次数的期望。

10. 一罐内有 a 个白球及 b 个黑球。从中任取一球，如果取出白球则将它放回去；如果是黑球，则从另一罐内拿一白球替换它放回去。在重复 n 次这样的作法后，该罐内白球数的期望记为 M_n。

(a) 导出递推公式

$$M_{n+1} = \left(1 - \frac{1}{a+b}\right) M_n + 1$$

(b) 利用(a)证明

$$M_n = a + b - b\left(1 - \frac{1}{a+b}\right)^n$$

(c) 第$(n+1)$次取出白球的概率是多少？

11. 连续的随机装填问题。考虑区间$(0,x)$且假设在此区间内我们装填随机单位区间（它们的左端点都服从$(0,x-1)$上的均匀分布）的方法如下。设第一个随机区间是 I_1。若 $I_1,\cdots,I_k$ 已装进此区间，下一个随机单位区间如果与区间 $I_i,\cdots,I_k$ 都不相交，将被装填进去，并记为 I_{k+1}。如果它与 $I_1,\cdots,I_k$ 中之任一相交时，我们就去掉它，再查看下一个随机区间。这种作法一直不断地进行到没有余地再装进一个单位区间为止（即已装进去的区间之间的间隔全部小于 1）。令 $N(x)$表示在$[0,x]$中按此法装进去的单位区间的个数。

例如，设 $x=5$，且相继的随机区间是$(0.5,1.5)$，$(3.1,4.1)$，$(4,5)$，$(1.7,2.7)$，则$N(5)=3$ 且装法如图 1.9.1 所示。记$M(x)=E[N(x)]$。证明：M 满足

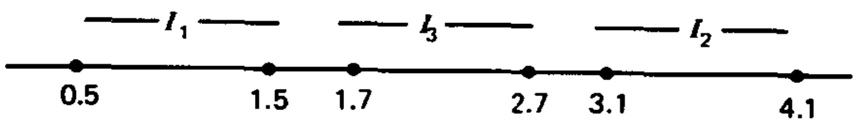


图 1.9.1

$$M(x) = 0, \quad x < 1$$

$$M(x) = \frac{2}{x-1}\int_0^{x-1} M(y)\,dy + 1, \quad x > 1$$

12. 设 $U_1,U_2,\cdots$ 是独立的$(0,1)$上均匀分布的随机变量，且以 N 记满足下式

的 $n, n\geqslant 0$,之最小值

$$\prod_{i=1}^{n} U_i \geqslant e^{-\lambda} > \prod_{i=1}^{n+1} U_i, \qquad 其中 \prod_{i=1}^{0} U_i \equiv 1$$

证明:N 是均值为 λ 的泊松随机变量。(提示:按 n 归纳,并对 U_1 取条件去证 $P\{N=n\}=e^{-\lambda}\lambda^n/n!$。)

13. 在给定 Y 的条件下,X 的条件方差定义为

$$\mathrm{Var}(X|Y) = E[(X - E[X|Y])^2/Y]$$

证明条件方差公式,即

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y])$$

用此式去求例 1.5 (b) 中的 Var (X),且通过对矩母函数求导来检验你的结果。

14. 在选票问题中计算 A 在计票过程中不曾落后的概率。

15. 考虑一个赌徒,他在每次赌局中分别以概率 p 及 $(1-p)$ 赢得一元或输掉一元。若赌徒开始有 n 元,问他输光之前恰好赌了 $n+2i$ 局的概率是多少?(提示:利用选票定理。)

16. 验证指数随机变量的均值与方差的公式。

17. 如果 X_1, X_2, …, X_n 是独立同分布的指数变量,参数为 λ。证明:$\sum_{i=1}^{n} X_i$ 具有参数为 (n, λ) 的 Γ-分布,即证明 $\sum_{i=1}^{n} X_i$ 的密度函数为

$$f(t) = \lambda e^{-\lambda t}(\lambda t)^{n-1}/(n-1)!, \quad t \geqslant 0$$

18. 在例 1.6 (a) 中如果营业员 i 的服务时间是参数为 λ_i 的指数变量,$i=1$, 2。计算 A 最后离去的概率。

19. 如果 X 与 Y 是独立指数随机变量,均值分别为 $1/\lambda_1$, $1/\lambda_2$,计算 $Z=\min(X, Y)$ 的分布。在 $Z=X$ 的条件下,Z 的条件分布是什么?

20. 证明函数方程

$$g(s+t) = g(s) + g(t)$$

的唯一连续解是 $g(s)=cs$。

21. 对于任意的连续分布 F,导出第 i 个记录值的分布(见例 1.6(b))。

22. 若 X_1 与 X_2 是独立的非负连续随机变量,证明:

$$P\{X_1 < X_2 | \min(X_1, X_2) = t\} = \frac{\lambda_1(t)}{\lambda_1(t) + \lambda_2(t)}$$

其中 $\lambda_i(t)$ 是 X_i 的失效率函数。

参考文献

文献5与8是概率论及其应用的初等入门书。文献1，2，3与4给出概率论与随机过程的严格测度论处理方法。例1.5(f)取自文献7，而例1.6(b)取自文献6。

1. P. Billingsley, *Probability and Measure*, Wiley, New York, 1979.
2. L. Breiman, *Probability*, Addison—Nesley, Reading, Mass., 1968.
3. K. L. Chung, *A Course in Probability Theory*, 2nd ed., Academic Press, New York, 1974.
   (钟开莱，概率论教程，刘文、吴让泉译，上海科技出版社，1989。)
4. J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
5. W. Feller, *An Introduction to Probability Theory and It's Applications*, Vol. 1, Wiley, New York, 1957.
   (费勤，概率论及其应用，科学出版社，上册，林向清、胡迪鹤译，1964；下册，刘文译，1979。)
6. M. Neuts, "Waiting Times Between Record Observations", Journal of Applied Probability, 7 (1967), pp. 206—208.
7. G. F. Page, "Vacancies on a Line", Journal of the Royal Statistic Society, Series B, 21, No. 2 (1959), pp. 364—370.
8. S. Ross, *A First Course in Probability*, Macmillan, New York, 1967.
   (罗斯，概率论初级教程，李漳南、杨振明译，人民教育出版社，1982。)

2
泊松过程

2.1 泊松过程

随机过程 $\{N(t),t\geqslant 0\}$ 称为一个计数过程，若 $N(t)$ 表示到时刻 t 为止已发生的“事件”的总数。因此一个计数过程 $N(t)$ 必须满足：

(i) $N(t)\geqslant 0$，

(ii) $N(t)$ 是整数值，

(iii) 若 $s<t$，则 $N(s)\leqslant N(t)$，

(iv) 当 $s<t$ 时，$N(t)-N(s)$ 等于区间 $(s,t]$ 中发生的事件的个数。

如果在不相交的时间区间中发生的事件个数是独立的，则称计数过程有独立增量。例如，这意味着到时刻 t 已发生的事件个数（即 $N(t)$）必须独立于时刻 t 与 $t+s$ 之间所发生的事件数（即 $N(t+s)-N(t)$）。

若在任一时间区间中发生的事件个数的分布只依赖于时间区间的长度，则称计数过程有平稳增量。换言之，若对一切 $t_1<t_2$ 及 $s>0$，在区间 $(t_1+s,t_2+s]$ 中事件的个数（即 $N(t_2+s)-N(t_1+s)$）与区间 $(t_1,t_2]$ 中事件的个数（即 $N(t_2)-N(t_1)$）有相同的分布，则

过程有平稳增量。

泊松过程是计数过程的最重要的类型之一，其定义如下。

定义 2.1.1　计数过程 $\{N(t), t\geqslant 0\}$ 称为泊松过程，具有参数 λ，$\lambda>0$，如果

(i) $N(0)=0$，

(ii) 过程有独立增量，

(iii) 在任一长度为 t 的区间中事件的个数服从均值为 λt 的泊松分布。即对一切 s，$t\geqslant 0$，

$$P\{N(t+s)-N(s)=n\}=e^{-\lambda t}\frac{(\lambda t)^n}{n!},\qquad n=0,1,\cdots$$

注意，从条件(iii)可知泊松过程有平稳增量且 $E[N(t)]=\lambda t$，这正是称 λ 为此过程的速率或强度的原因（单位时间内发生的事件的平均个数）。

为了确定一个任意的计数过程实际上是一泊松过程，我们必须证明它满足条件(i),(ii)及(iii)。条件(i)只是说明事件的计数是从时刻 $t=0$ 开始的。条件 (ii) 通常可从我们对过程了解的情况去直接验证。然而全然不清楚如何去确定条件 (iii) 是否满足。为此泊松过程的一个等价定义将是有用的。

作为给出泊松过程的第二个定义的准备，先定义一个函数 f 是 $o(h)$ 的概念。

定义　若

$$\lim_{h\to 0}\frac{f(h)}{h}=0$$

则称函数 f 是 $o(h)$。

现在我们能给出泊松过程的另一个定义。

定义 2.1.2　计数过程 $\{N(t), t\geqslant 0\}$ 称为泊松过程，参数为 λ，$\lambda>0$，如果

(i) $N(0)=0$，

(ii) 过程有平稳与独立增量，

(iii) $P\{N(h)=1\}=\lambda h+o(h)$,

(Ⅳ) $P\{N(h)\geqslant 2\}=o(h)$。

定理 2.1.3 定义 2.1.1 与 2.1.2 是等价的。

证明 首先我们证明定义 2.1.2 蕴含定义 2.1.1。为此设

$$P_n(t) = P\{N(t) = n\}$$

按以下方法导出一个关于 $P_0(t)$ 的微分方程:

$$\begin{aligned} P_0(t + h) &= P\{N(t + h) = 0\} \\ &= P\{N(t) = 0, N(t + h) - N(t) = 0\} \\ &= P\{N(t) = 0\}P\{N(t + h) - N(t) = 0\} \\ &= P_0(t)[1 - \lambda h + o(h)] \end{aligned}$$

其中最后两个等式由假定(ii)与(iii)及(iv)蕴含了 $P\{N(h)=0\}=1-\lambda h+o(h)$ 这一事实而得到。因此

$$\frac{P_0(t + h) - P_0(t)}{h} = - \lambda P_0(t) + \frac{o(h)}{h}$$

令 $h \to 0$ 得

$$P'_0(t) = - \lambda P_0(t)$$

或

$$\frac{P'_0(t)}{P_0(t)} = - \lambda$$

经积分得

$$\log P_0(t) = - \lambda t + c$$

或

$$P_0(t) = Ke^{-\lambda t}$$

既然 $P_0(0) = P\{N(0) = 0\} = 1$，得到

(2.1.1) $$P_0(t) = e^{-\lambda t}$$

类似地，当 $n \geqslant 1$ 时

$$\begin{aligned} P_n(t + h) &= P\{N(t + h) = n\} \\ &= P\{N(t) = n, N(t + h) - N(t) = 0\} \\ &\quad + P\{N(t) = n - 1, N(t + h) - N(t) = 1\} \\ &\quad + P\{N(t + h) = n, N(t + h) - N(t) \geqslant 2\} \end{aligned}$$

然而，由(iv)，上式最后一项是 $o(h)$；因而，利用(ii)得

$$P_n(t + h) = P_n(t)P_0(h) + P_{n-1}(t)P_1(h) + o(h)$$

$$= (1 - \lambda h)P_n(t) + \lambda h P_{n-1}(t) + o(h)$$

于是

$$\frac{P_n(t+h) - P_n(t)}{h} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(h)}{h}$$

令 $h \to 0$

$$P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t)$$

或等价地，

$$e^{\lambda t}[P_n'(t) + \lambda P_n(t)] = \lambda e^{\lambda t} P_{n-1}(t)$$

因此，

(2.1.2)
$$\frac{d}{dt}(e^{\lambda t} P_n(t)) = \lambda e^{\lambda t} P_{n-1}(t)$$

现在由(2.1.1)，当 $n=1$ 时我们有

$$\frac{d}{dt}(e^{t\lambda} P_1(t)) = \lambda$$

或

$$P_1(t) = (\lambda t + c)e^{-\lambda t}$$

又因 $P_1(0)=0$,得

$$P_1(t) = \lambda t e^{-\lambda t}$$

为证明 $P_n(t)=e^{-\lambda t}(\lambda t)^n/n!$,我们用数学归纳法,因此先假定$(n-1)$时它成立。由(2.1.2)

$$\frac{d}{dt}(e^{\lambda t} P_n(t)) = \frac{\lambda (\lambda t)^{n-1}}{(n-1)!}$$

从而

$$e^{\lambda t} P_n(t) = \frac{(\lambda t)^n}{n!} + c$$

因 $P_n(0)=P\{N(0)=n\}=0$,得

$$P_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

于是定义 2.1.2 蕴含了定义 2.1.1。逆命题的证明留给读者去作。

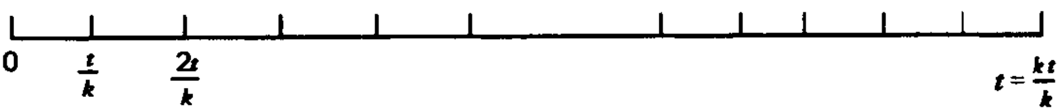


图 2.1.1

注记　$N(t)$有泊松分布这个结果是二项分布的泊松逼近的推论。为明瞭这点，把区间$[0,t]$划分为k个相等的部分，其中k非常大（图 2.1.1）。首先注意，在任一子区间中发生两个或更多个事件的概率当$k\to\infty$时趋于 0，这可从下面的推导得知。

$$\begin{aligned}&P\{\text{在任一子区间中有 2 个或更多个事件}\}\\&\quad\leqslant\sum_{i=1}^{k}P\{\text{在第 } i \text{ 个子区间中有 2 个或更多个事件}\}\\&\quad=ko(t/k)\\&\quad=t\,\frac{o(t/k)}{t/k}\\&\quad\to 0,\ \text{当 } k\to\infty\end{aligned}$$

因此,$N(t)$将(以趋于 1 的概率)正好等于有一个事件发生的子区间的个数。然而由于有平稳独立增量，此个数有参数为k及$p=\lambda t/k+o(t/k)$的二项分布。因此由二项分布的泊松逼近得知，令k趋于∞，$N(t)$将有泊松分布,其均值为

$$\lim_{k\to\infty}k\left[\lambda\frac{t}{k}+o\left(\frac{t}{k}\right)\right]=\lambda t+\lim_{k\to\infty}\left[t\cdot\frac{o(t/k)}{t/k}\right]=\lambda t$$

2.2　来到间隔与等待时间的分布

考虑一泊松过程，以X_1记第一个事件来到的时刻。对$n\geqslant1$，以X_n记第$(n-1)$个到第n个事件之间的时间。序列$\{X_n,n\geqslant1\}$称为来到间隔序列。

现在来确定X_n的分布。为此首先注意到事件$\{X_1>t\}$发生当且仅当泊松过程在区间$[0,t]$内没有事件发生，因而

$$P\{X_1>t\}=P\{N(t)=0\}=e^{-\lambda t}$$

因此,X_1具有均值为$1/\lambda$的指数分布。求已知X_1的条件下X_2的分布。然而

$$\begin{aligned}P\{X_2>t\,|\,X_1=s\}&=P\{\text{在}(s,s+t]\text{ 内没有事件 }|\,X_1=s\}\\&=P\{\text{在}(s,s+t]\text{ 内没有事件}\}\quad\text{（由独立增量）}\end{aligned}$$

$$= e^{-\lambda t} \quad (\text{由平稳增量})$$

所以，从上可得，X_2 也是一个具有均值 $1/\lambda$ 的指数随机变量，且 X_2 独立于 X_1。重复同样的推导得下列命题。

命题 2.2.1

X_n，$n=1$，2，…，为独立同分布的均值为 $1/\lambda$ 的指数随机变量。

注记　这个命题不应使我们惊奇。平稳独立增量的假定等价于说在概率意义上过程在任何时刻都重新开始，即从任何时刻起过程独立于先前已发生的一切（由独立增量），且有与原过程完全一样的分布（由平稳增量）。换言之，过程无记忆，因此指数间隔是预料之中的。

另一个感兴趣的量是 S_n，第 n 个事件来到的时间，也称为第 n 个事件的等待时间。因

$$S_n = \sum_{i=1}^{n} X_i, \qquad n \geqslant 1$$

用矩母函数容易证明，命题 2.2.1 蕴含了 S_n 有参数为 n 与 λ 的 Γ-分布，即其概率密度为

$$f(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \qquad t \geqslant 0$$

上式也可用下述方式导出。注意到第 n 个事件在时刻 t 或之前发生当且仅当到时间 t 已发生的事件数目至少是 n，即

$$N(t) \geqslant n \iff S_n \leqslant t$$

因此，

$$P\{S_n \leqslant t\} = P\{N(t) \geqslant n\} = \sum_{j=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!}$$

对上式求导，得 S_n 的密度函数是

$$f(t) = -\sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^j}{j!} + \sum_{j=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^{j-1}}{(j-1)!}$$

$$= \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}$$

命题 2.2.1 又给我们定义泊松过程的另一个方法。我们从一列均值为 $1/\lambda$ 的独立同分布的指数随机变量 $\{X_n, n \geqslant 1\}$ 出发。现在说定第 n 个事件在时刻 S_n 发生，这里

$$S_n \equiv X_1 + X_2 + \cdots + X_n$$

这样就定义了一个计数过程，且所得计数过程 $\{N(t), t \geqslant 0\}$ 就是参数为 λ 的泊松过程。

2.3 来到时刻的条件分布

假设已知到时间 t 泊松过程恰发生了一个事件，我们要确定这一事件发生的时刻的分布。因为泊松过程有平稳独立增量，看来有理由认为 $[0, t]$ 内长度相等的区间包含这个事件的概率应该相同。换言之，这个事件的来到时刻应在 $[0, t]$ 上均匀分布。容易验证此事，因为对 $s \leqslant t$ 有

$$P\{X_1 < s \mid N(t) = 1\} = \frac{P\{X_1 < s, N(t) = 1\}}{P\{N(t) = 1\}}$$

$$= \frac{P\{\text{在}[0,s)\text{ 内有 1 个事件,在}[s,t]\text{ 内没有事件}\}}{P\{N(t) = 1\}}$$

$$= \frac{P\{\text{在}[0,s)\text{ 内有 1 个事件}\}P\{\text{在}[s,t]\text{ 内没有事件}\}}{P\{N(t) = 1\}}$$

$$= \frac{\lambda s e^{-\lambda s} e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}}$$

$$= \frac{s}{t}$$

可以推广这个结果，但为此需先引进顺序统计量的概念。

设 $Y_1, Y_2, \cdots, Y_n$ 是 n 个随机变量，如果 $Y_{(k)}$ 是 $Y_1, \cdots Y_n$，中的第 k 个最小值，$k=1, 2, \cdots, n$，则称 $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}$ 是对应于 $Y_1, Y_2, \cdots, Y_n$ 的顺序统计量。若 Y_i，$i=1, 2, \cdots, n$，是独立同分布的连续随机变量，具有概率密度 f，则顺序统计量 $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}$ 的联合密度为

$$f(y_1, y_2, \cdots, y_n) = n! \prod_{i=1}^{n} f(y_i), \quad y_1 < y_2 < \cdots < y_n$$

上式成立是由于以下原因：(i)（$Y_{(1)}$，…，$Y_{(n)}$）将等于（y_1，…，y_n），如果（Y_1，Y_2，…，Y_n）等于（y_1，…，y_n）的 n! 个排列中的任一个；(ii) 当（y_{i_1}，y_{i_2}，…，y_{i_n}）是（y_1，…，y_n）的一个排列时，（Y_1，Y_2，…，Y_n）等于 y_{i_1}，y_{i_2}，…，y_{i_n}的概率密度是$f(y_{i_1}) \cdots f(y_{i_n}) = \prod_{i=1}^{n} f(y_i)$。

若 Y_i，$i=1$，2，…，n，都在$(0,t)$上均匀分布，则由上面的讨论可知，顺序统计量 $Y_{(1)}$，$Y_{(2)}$，…，$Y_{(n)}$的联合密度函数是

$$f(y_1, \cdots, y_n) = \frac{n!}{t^n}, \qquad 0 < y_1 < y_2 \cdots < y_n < t$$

现在已为下列很有用的定理作好了准备。

定理 2.3.1 在已知 $N(t)=n$ 的条件下，n 个来到时刻 S_1，…，S_n 与相应于 n 个$(0,t)$上均匀分布的独立随机变量的顺序统计量有相同的分布。

证明 我们来计算给定 $N(t)=n$ 时，$S_1, \cdots, S_n$ 的条件密度函数，设 $0<t_1<t_2<\cdots<t_{n+1}=t$，且取 h_i 充分小使得 $t_i+h_i<t_{i+1}$，$i=1,2,\cdots,n$。现在

$P\{t_i \leqslant S_i \leqslant t_i + h_i, i = 1,2,\cdots,n \mid N(t) = n\}$

$$= \frac{P\{\text{在}[t_i, t_i + h_i]\text{ 中恰有一个事件}, i = 1,2,\cdots,n, \text{在}[0,t]\text{ 的别处无任何事件}\}}{P\{N(t) - n\}}$$

$$= \frac{\lambda h_1 e^{-\lambda h_1} \cdots \lambda h_n e^{-\lambda h_n} e^{-\lambda(t-h_1-h_2\cdots-h_n)}}{e^{-\lambda t}(\lambda t)^n/n!}$$

$$= \frac{n!}{t^n} h_1 h_2 \cdots h_n$$

因此

$$\frac{P\{t_i \leqslant S_i \leqslant t_i + h_i, i = 1,2,\cdots,n \mid N(t) = n\}}{h_1 h_2 \cdots h_n} = \frac{n!}{t^n}$$

令 $h_i \to 0$，我们得到 S_1，S_2，…，S_n 在已知 $N(t)=n$ 的条件下的条件密度为

$$f(t_1, \cdots, t_n) = \frac{n!}{t^n}, \quad 0 < t_1 < \cdots < t_n$$

证毕。

注记　直观地，我们常说在已知$(0, t]$内发生n个事件的条件下，各事件发生的时刻$S_1, \cdots, S_n$，看作不排顺序的随机变量，是相互独立的且服从$(0,t)$上的均匀分布。

例 2.3(a)　假设乘客按照参数为λ的泊松过程来到一个火车站。若火车在时刻t启程，让我们计算在$(0,t)$内到达的乘客的等待时间的总和的期望，即求$E\left[\sum_{i=1}^{N(t)}(t-S_i)\right]$，其中$S_i$是第$i$个乘客的来到时刻。对$N(t)$取条件得

$$E\left[\sum_{i=1}^{N(t)}(t-S_i)\,|\,N(t)=n\right]=E\left[\sum_{i=1}^{n}(t-S_i)\,|\,N(t)=n\right]$$

$$=nt-E\left[\sum_{i=1}^{n}S_i\,|\,N(t)=n\right]$$

现在以$U_1, U_2, \cdots, U_n$记n个独立的$(0,t)$上均匀分布的随机变量，则

$$E\left[\sum_{i=1}^{n}S_i\,|\,N(t)=n\right]=E\left[\sum_{i=1}^{n}U_{(i)}\right]\quad(\text{由定理 } 2.3.1)$$

$$=E\left(\sum_{i=1}^{n}U_i\right)\quad\left(\text{因}\sum_{i=1}^{n}U_{(i)}=\sum_{i=1}^{n}U_i\right)$$

$$=\frac{nt}{2}$$

因此

$$E\left[\sum_{i=1}^{N(t)}(t-S_i)\,|\,N(t)=n\right]=nt-\frac{nt}{2}=\frac{nt}{2}$$

及

$$E\left[\sum_{i=1}^{N(t)}(t-S_i)\right]=\frac{t}{2}E[N(t)]=\frac{1}{2}\lambda t^2$$

作为定理 2.3.1 的一个重要应用，假设参数为λ的泊松过程的各个事件被分成 I-型或 II-型的，且假设一个事件被归作 I-型的概率依赖于它发生的时刻。具体地说，假设一事件在时刻s发生，以概率$P(s)$被归作 I-型，而以概率$1-P(s)$归作 II-型，且与其它事件归作什么类型相互独立。利用定理 2.3.1 可以证明下面的命题。

命题 2.3.2

若$N_i(t)$表示到时刻t，i-型事件发生的个数，$i=1,2$，则$N_1(t)$

与 $N_2(t)$是独立的泊松随机变量，分别有均值 λtp 及 $\lambda t(1-p)$. 其中

$$p=\frac{1}{t}\int_0^t P(s)ds$$

证明 对 $N(t)$取条件，我们计算 $N_1(t)$与 $N_2(t)$的联合分布：

$$\begin{aligned}
&P\{N_1(t)=n,N_2(t)=m\}\\
&=\sum_{k=0}^{\infty}P\{N_1(t)=n,N_2(t)=m|N(t)=k\}P\{N(t)=k\}\\
&=P\{N_1(t)=n,N_2(t)=m|N(t)=n+m\}P\{N(t)=n+m\}
\end{aligned}$$

现在考虑在区间$[0,t]$中发生的任一事件。如果它在时刻 s 发生，则它是 I-型事件的概率为 $P(s)$。而由定理 2.3.1 此事件发生的时刻在$(0,t)$上均匀分布，所以，它是一个 I-型事件的概率为

$$p=\frac{1}{t}\int_0^t P(s)ds$$

与其它事件为什么类型相互独立。因此，$P\{N_1(t)=n,N_2(t)=m|N(t)=n+m\}$刚好等于 $n+m$ 次试验中 n 次成功 m 次失败的概率，而 p 是各次试验成功的概率，即

$$P\{N_1(t)=n,N_2(t)=m|N(t)=n+m\}=\binom{n+m}{n}p^n(1-p)^m$$

从而

$$\begin{aligned}
&P\{N_1(t)=n,N_2(t)=m\}\\
&\qquad=\frac{(n+m)!}{n!m!}p^n(1-p)^m\frac{(\lambda t)^{n+m}}{(n+m)!}e^{-\lambda t}\\
&\qquad=e^{-\lambda tp}\frac{(\lambda tp)^n}{n!}e^{-\lambda t(1-p)}\frac{(\lambda t(1-p))^m}{m!}
\end{aligned}$$

证毕。

上述命题的重要性可由下例来说明。

例 2.3(b)　无穷个服务员的泊松排队系统。设顾客按照参数为 λ 的泊松过程到达一服务站。顾客一到，便立即由无穷个服务员中的一个提供服务，且假设服务时间是独立的，有共同的分布 G。

为计算在时刻 t 服务完毕的顾客数与尚在服务中的顾客数的联合分布，我们称一顾客为 I-型的，到时刻 t 他已服务完毕；称一顾客是 II-型的，他在时刻 t 尚未服务完毕。现在，如果顾客在时刻 s（$s\leqslant t$）来到，那么当他的服

务时间小于 $t-s$ 时，他是 I-型的，又因服务时间的分布是 G，这个概率是 $G(t-s)$。因此

$$P(s)=G(t-s),\quad s\leqslant t$$

于是从命题 2.3.2 得到，$N_1(t)$（已服务完毕的顾客数）的分布是均值为

$$E[N_1(t)]=\lambda\int_0^t G(t-s)ds=\lambda\int_0^t G(y)dy$$

的泊松分布。类似地 $N_2(t)$（时刻 t 尚在服务中的顾客数）有泊松分布，其均值为

$$E[N_2(t)]=\lambda\int_0^t \overline{G}(y)dy$$

且 $N_1(t)$ 与 $N_2(t)$ 独立。

下例进一步说明定理 2.3.1 的应用。

例 2.3(c) 假设一部仪器承受到冲击，冲击遵循参数为 λ 的泊松过程来到。第 i 次冲击造成损失 D_i。假定 $D_i, i\geqslant 1$，独立同分布且与 $\{N(t), t\geqslant 0\}$ 独立，$N(t)$ 表示 $[0,t]$ 中的冲击次数。假定冲击引起的损伤随时间而指数地衰减，即若一个冲击造成的初始损伤为 D，时间 t 之后它造成的损伤则是 $De^{-\alpha t}$，$\alpha>0$。

若假定损伤是可加的，则在 t 时的损伤 $D(t)$ 可表示为

$$D(t)=\sum_{i=1}^{N(t)}D_ie^{-\alpha(t-S_i)}$$

其中 S_i 表示第 i 次冲击来到的时刻。我们能求得 $E[D(t)]$ 如下：

$$\begin{aligned}
E[D(t)\mid N(t)=n]&=E\Big[\sum_{i=1}^{N(t)}D_ie^{-\alpha(t-S_i)}\mid N(t)=n\Big]\\
&=E\Big[\sum_{i=1}^{n}D_ie^{-\alpha(t-S_i)}\mid N(t)=n\Big]\\
&=\sum_{i=1}^{n}E[D_ie^{-\alpha(t-S_i)}\mid N(t)=n]\\
&=\sum_{i=1}^{n}E[D_i\mid N(t)=n]E[e^{-\alpha(t-S_i)}\mid N(t)=n]\\
&=E[D]\sum_{i=1}^{N}E[e^{-\alpha(t-S_i)}\mid N(t)=n]\\
&=E[D]E\Big[\sum_{i=1}^{n}e^{-\alpha(t-S_i)}\mid N(t)=n\Big]\\
&=E[D]e^{-\alpha t}E\Big[\sum_{i=1}^{n}e^{\alpha S_i}\mid N(t)=n\Big]
\end{aligned}$$

今设$U_1, U_2, \cdots, U_n$是独立同分布的$[0,t]$上的均匀分布随机变量，则由定理 2.3.1 有

$$E\left[\sum_{i=1}^{n} e^{\alpha S_i} \mid N(t) = n\right] = E\left[\sum_{i=1}^{n} e^{\alpha U_{(i)}}\right]$$
$$= E\left[\sum_{i=1}^{n} e^{\alpha U_i}\right]$$
$$= \frac{n}{t}\int_0^t e^{\alpha x} dx$$
$$= \frac{n}{\alpha t}(e^{\alpha t} - 1)$$

因此，

$$E[D(t) \mid N(t)] = \frac{N(t)}{\alpha t}(1 - e^{-\alpha t})E[D]$$

取期望得

$$E[D(t)] = \frac{\lambda E(D)}{\alpha}(1 - e^{-\alpha t})$$

注记　得到$E[D(t)]$的另一途径是将区间$(0,t)$分成长为h的互不相交的区间，然后将这些区间内发生的冲击在时刻t造成的后果叠加。更确切地说，设给定h，且定义X_i是区间$I_i \equiv (ih, (i+1)h)$，$i=0, 1, \cdots, [t/h]$，内来到的一切冲击在时刻t的损伤之和，其中$[a]$表示小于或等于a的最大整数。我们有表达式

$$D(t) = \sum_{i=0}^{[t/h]} X_i$$

从而

$$E[D(t)] = \sum_{i=0}^{[t/h]} E[X_i]$$

为了计算$E[X_i]$，以区间I_i中是否有冲击来到为条件，可得

$$E[D(t)] = \sum_{i=0}^{[t/h]} \{\lambda h E[D e^{-\alpha(t-L_i)}] + o(h)\}$$

其中L_i是区间I_i中冲击来到的时刻。因此，

$$E[D(t)] = \lambda E[D] E\left[\sum_{i=0}^{[t/h]} h e^{-\alpha(t-L_i)}\right] + \left[\frac{t}{h}\right] o(h) \tag{2.3.1}$$

但因$L_i \in I_i$，令$h \to 0$可得

$$\sum_{i=0}^{[t/h]} he^{-\alpha(t-L_i)} \longrightarrow \int_0^t e^{-\alpha(t-y)} dy = \frac{1-e^{-\alpha t}}{\alpha}$$

于是在（2.3.1）中令$h \to 0$，有

$$E[D(t)] = \frac{\lambda E(D)}{\alpha}(1 - e^{-\alpha t})$$

值得注意的是以上所述正是以下推断的较为严格的叙述：因为在区间$(y, y+dy)$中发生一次冲击的概率为λdy，又因为它的损伤在时刻t等于$e^{-\alpha(t-y)}$乘以它的初始损伤，来自$(y, y+dy)$的冲击在时刻t的平均损伤为

$$\lambda dy E[D] e^{-\alpha(t-y)}$$

从而

$$E[D(t)] = \lambda E[D] \int_0^t e^{-\alpha(t-y)} dy$$
$$= \frac{\lambda E[D]}{\alpha} (1 - e^{-\alpha t})$$

2.3.1 $M|G|1$ 忙期

考虑称为$M|G|1$的排队系统，其中顾客按参数为λ的泊松过程来到。到达时，若服务员有空就接受服务，否则就排队。相继的服务时间是独立的且服从同一分布G，而且又独立于来到的过程。当一名来客发现服务员有空时，我们就说一个忙期开始，在系统中不再有顾客时它结束。我们要计算一个忙期的长度的分布。

假设一个忙期刚好在某个时刻开始，我们规定这个时刻为0，以S_k记又来到k个顾客的时刻。（于是，S_k服从参数为k，λ的Γ-分布。）又设Y_1，Y_2，…为服务时间序列。于是忙期持续时间为t且由n次服务组成当且仅当：

(i) $S_k \leqslant Y_1 + \cdots + Y_k$，　$k = 1, 2, \cdots, n-1$，

(ii) $Y_1 + \cdots + Y_n = t$，

(iii) 在$(0, t)$内有$n-1$个顾客来到。

(i)是必要的，因为如果$S_k > Y_1 + \cdots + Y_k$则在初始顾客之后来到的第k个顾客会发现系统中无顾客，从而在第$k+1$次服务之前

(于是在第 n 次服务之前)忙期就已结束了。(ii)和(iii)的推导是直接了当的，留给读者完成。

直观地作推导(把密度当作概率)，由上述讨论可见

(2.3.2) P\{忙期长为 t 且由 n 次服务组成\}

$$=P\{Y_1+\cdots+Y_n=t,\text{在}(0,t)\text{内 } n-1 \text{ 个顾客来到},\ S_k\leqslant Y_1+\cdots+Y_k,k=1,2,\cdots,n-1\}$$

$$=P\{S_k\leqslant Y_1+\cdots+Y_k,k=1,2,\cdots,n-1\mid \text{在}(0,t)\text{内 } n-1 \text{ 个顾客来到},Y_1+\cdots+Y_n=t\}\times P\{\text{在}(0,t)\text{内 } n-1 \text{ 个顾客来到},Y_1+\cdots+Y_n=t\}$$

因为来到过程独立于服务时间，于是

(2.3.3) P\{在$(0,t)$内 $n-1$ 个顾客来到，$Y_1+\cdots+Y_n=t$\}

$$=e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}dG_n(t)$$

其中 G_n 是 G 自身的 n 次卷积。此外，以定理 2.3.1 可知，在$(0,t)$内有 $n-1$ 个顾客来到的条件下，依次的来到时刻如同 $n-1$ 个独立的$(0,t)$上均匀分布的随机变量的顺序统计量。利用这个事实，连同(2.3.3)与(2.3.2)得

(2.3.4) P\{忙期长为 t 且由 n 次服务组成\}

$$=e^{-\lambda t}\frac{(\lambda t)^{n-1}}{(n-1)!}dG_n(t)\times P\{\tau_k\leqslant Y_1+\cdots+Y_k,\ k=1,2,\cdots,n-1\mid Y_1+\cdots+Y_n=t\}$$

其中 $\tau_1,\tau_2,\cdots,\tau_{n-1}$ 与 $\{Y_1,\cdots,Y_n\}$ 独立，且表示 $n-1$ 个$(0,1)$上均匀分布随机变量的顺序统计量。

为计算(2.3.4)中余下的概率需要一些引理。引理 2.3.3 是初等的，其证明留作一个习题。

引理 2.3.3

设 Y_1，Y_2，$\cdots$，Y_n 为独立同分布的非负随机变量，则

$$E[Y_1+\cdots+Y_k|Y_1+\cdots+Y_n=y]=\frac{k}{n}y,\quad k=1,2,\cdots,n$$

引理 2.3.4

设τ_1，τ_2，…，τ_n为n个独立的$(0,t)$上均匀分布随机变量的顺序统计量。设Y_1，Y_2，…是独立同分布的非负随机变量且独立于$\{\tau_1,\tau_2,\cdots,\tau_n\}$，则

$$(2.3.5)\quad P\{Y_1+\cdots+Y_k<\tau_k,k=1,2,\cdots,n|Y_1+\cdots+Y_n=y\}$$
$$=\begin{cases}1-y/t, & 0<y<t\\ 0, & \text{其它}\end{cases}$$

证明 用归纳法证。当$n=1$时，我们必须计算当τ_1在$(0,t)$上均匀分布时，$P\{Y_1<\tau_1|Y_1=y\}$之值。但是

$$P\{Y_1<\tau_1|Y_1=y\}=P\{y<\tau_1\}=1-y/t,\qquad 0<y<t$$

所以假定引理对$n-1$成立，考虑n的情形。因为对$y\geqslant t$结论是显然成立，所以假设$y<t$。为了利用归纳假设，我们以$Y_1+\cdots+Y_{n-1}$及τ_n的值为条件来计算(2.3.5)的左端，且利用下述结论：在$\tau_n=u$的条件下，$\tau_1,\cdots,\tau_{n-1}$与$n-1$个$(0,u)$上均匀分布变量的顺序统计量同分布(见习题13)。按此做法，对$s<y$有

$$(2.3.6)\quad P\{Y_1+\cdots+Y_k<\tau_k,k=1,\cdots,n|Y_1+\cdots+Y_{n-1}=s,$$
$$\tau_n=u,Y_1+\cdots+Y_n=y\}$$
$$=\begin{cases}P\{Y_1+\cdots+Y_k<\tau_k^*,k=1,\cdots,n-1|Y_1+\cdots+Y_n=s\}, & y<u\\ 0, & y\geqslant u\end{cases}$$

其中$\tau_1^*,\cdots,\tau_{n-1}^*$是$n-1$个独立的$(0,u)$上均匀分布随机变量的顺序统计量。由归纳法假设可见(2.3.6)的右边等于

$$\text{右边}=\begin{cases}1-s/u, & y<u\\ 0, & \text{其它}\end{cases}$$

因此，对$y<u$，

$$P\{Y_1+\cdots+Y_k<\tau_k,k=1,2,\cdots,n|Y_1+\cdots+Y_{n-1},$$
$$\tau_n=u,Y_1+\cdots+Y_n=y\}$$
$$=1-\frac{Y_1+\cdots+Y_{n-1}}{\tau_n}$$

从而，对$y<u$，

$$P\{Y_1+\cdots+Y_k<\tau_k,k=1,2,\cdots,n|\tau_n=u,Y_1+\cdots+Y_n=y\}$$
$$=E\left[1-\frac{Y_1+\cdots+Y_{n-1}}{\tau_n}\middle|\tau_n=u,Y_1+\cdots+Y_n=y\right]$$
$$=1-\frac{1}{u}E[Y_1+\cdots+Y_{n-1}|Y_1+\cdots+Y_n=y]$$

$$=1-\frac{n-1}{n}\frac{y}{u}$$

这里我们利用了上面的引理 2.3.3，再次取期望得

(2.3.7)
$$P\{Y_1+\cdots+Y_k<\tau_k, k=1,\cdots,n \mid Y_1+\cdots+Y_n=y\}$$
$$=E\left[1-\frac{n-1}{n}\frac{y}{\tau_n}\,\middle|\, y<\tau_n\right]P\{y<\tau_n\}$$
$$=P\{y<\tau_n\}-\frac{n-1}{n}yE\left[\frac{1}{\tau_n}\,\middle|\, y<\tau_n\right]P\{y<\tau_n\}$$

而 τ_n 的分布为

$$P\{\tau_n<x\}=P\{\max_{1\leqslant i\leqslant n}U_i<x\}$$
$$=P\{U_i<x, i=1,\cdots,n\}$$
$$=(x/t)^n,\quad 0<x<t$$

其中 U_i，$i=1$，…，n，是独立的 $(0,t)$ 上均匀分布变量。因此它的密度为

$$f_{\tau_n}(x)=\frac{n}{t}\left(\frac{x}{t}\right)^{n-1},\quad 0<x<t$$

于是

(2.3.8)
$$E\left[\frac{1}{\tau_n}\,\middle|\,\tau_n>y\right]P\{\tau_n>y\}=\int_y^t\frac{1}{x}\frac{n}{t}\left(\frac{x}{t}\right)^{n-1}dx$$
$$=\frac{n}{n-1}\left(\frac{t^{n-1}-y^{n-1}}{t^n}\right)$$

从(2.3.7)及(2.3.8)得，$y<t$ 时

$$P\{Y_1+\cdots+Y_k<\tau_k, k=1,2,\cdots,n \mid Y_1+\cdots+Y_n=y\}$$
$$=1-\left(\frac{y}{t}\right)^n-\frac{y\ (t^{n-1}-y^{n-1})}{t^n}$$
$$=1-y/t$$

证毕。

在回到忙期问题之前，我们还需要一个引理。

引理 2.3.5

设 τ_1，…，τ_{n-1} 为 $n-1$ 个独立的 $(0,t)$ 上均匀分布随机变量的顺序统计量，Y_1，Y_2，…是独立同分布的非负的随机变量，且与 $\{\tau_1, \cdots, \tau_{n-1}\}$ 独立，则

$$P\{Y_1+\cdots+Y_k<\tau_k, k=1,2,\cdots,n-1 \mid Y_1+\cdots+Y_n=t\}$$
$$=\frac{1}{n}$$

证明 为计算上述概率，我们利用引理 2.3.4，并对 $Y_1+\cdots+Y_{n-1}$ 取

条件。于是由引理 2.3.4，

$$P\{Y_1+\cdots+Y_k<\tau_k,k=1,\cdots,n-1|Y_1+\cdots+Y_{n-1}=y,\ Y_1+\cdots+Y_n=t\}$$

$$=P\{Y_1+\cdots+Y_k<\tau_k,\ k=1,\ \cdots,\ n-1|Y_1+\cdots+Y_{n-1}=y\}$$

$$=\begin{cases}1-y/t, & 0<y<t\\ 0\ , & \text{其它}\end{cases}$$

因为，$Y_1+\cdots+Y_{n-1}\leqslant Y_1+\cdots+Y_n$，我们有

$$P\{Y_1+\cdots+Y_k<\tau_k,k=1,\cdots,n-1|Y_1+\cdots+Y_n=t\}$$

$$=E\left[1-\frac{Y_1+Y_2+\cdots+Y_{n-1}}{t}\middle|Y_1+\cdots+Y_n=t\right]$$

$$=1-\frac{n-1}{n}\quad\text{（由引理 2.3.3）}$$

结果得证。

现在回到忙期长度与受到服务的顾客数的联合分布上来。由(2.3.4)，我们必须计算

$$P\{\tau_k\leqslant Y_1+\cdots+Y_k,k=1,\cdots,n-1|Y_1+\cdots+Y_n=t\}$$

只要 U 是$(0,t)$上的均匀分布随机变量，$t-U$ 也是如此，从而 τ_1，…，τ_{n-1}与 $t-\tau_{n-1}$，…，$t-\tau_1$ 有相同的联合分布，因此将 τ_k 全部换成 $t-\tau_{n-k}$，$1\leqslant k\leqslant n-1$，我们得到

$$P\{\tau_k\leqslant Y_1+\cdots+Y_k,k=1,\cdots,n-1|Y_1+\cdots+Y_n=t\}$$

$$=P\{t-\tau_{n-k}\leqslant Y_1+\cdots+Y_k,k=1,\cdots,n-1|Y_1+\cdots+Y_n=t\}$$

$$=P\{t-\tau_{n-k}\leqslant t-(Y_{k+1}+\cdots+Y_n),k=1,\cdots,n-1|Y_1+\cdots+Y_n=t\}$$

$$=P\{\tau_{n-k}\geqslant Y_{k+1}+\cdots+Y_n,k=1,\cdots,n-1|Y_1+\cdots+Y_n=t\}$$

$$=P\{\tau_{n-k}\geqslant Y_{n-k}+\cdots+Y_1,k=1,\cdots,n-1|Y_1+\cdots+Y_n=t\}$$

其中最后的等式成立是因为 Y_1，…，Y_n 与 Y_n，…，Y_1 有相同的联合分布，从而有关诸 Y_i 的任何概率命题用 Y_n 换 Y_1，Y_{n-1}换 Y_2，…，Y_{n-k+1}换 Y_k，…，Y_1 换 Y_n 后仍然成立。由此可见

$$P\{\tau_k\geqslant Y_1+\cdots+Y_k,k=1,\cdots,n-1|Y_1+\cdots+Y_n=t\}$$

$$=P\{\tau_k\geqslant Y_1+\cdots+Y_k,k=1,\cdots,n-1|Y_1+\cdots+Y_n=t\}$$

$$=1/n\quad\text{（由引理 2.3.5）}$$

因此，若令

$$B(t,n) = P\{\text{忙期的长} \leqslant t,\text{忙期中 } n \text{ 个顾客受到服务}\}$$

则从(2.3.4)得

$$\frac{d}{dt}B(t,n) = e^{-\lambda t}\frac{(\lambda t)^{n-1}}{n!}dG_n(t)$$

或

$$B(t,n) = \int_0^t e^{-\lambda t}\frac{(\lambda t)^{n-1}}{n!}dG_n(t)$$

忙期长度的分布，记作 $B(t)=\sum_{n=1}^{\infty}B(t,\ n)$，为

$$B(t) = \sum_{n=1}^{\infty}\int_0^t e^{-\lambda t}\frac{(\lambda t)^{n-1}}{n!}dG_n(t)$$

2.4 非齐次泊松过程

本节中我们要推广泊松过程，允许时刻 t 的来到速率(或强度)是 t 的函数。

定义 2.4.1 计数过程 $\{N(t),t\geqslant 0\}$ 称为非平稳或非齐次泊松过程，有强度函数 $\lambda(t),t\geqslant 0$，如果

(i) $N(0)=0$,

(ii) $\{N(t),t\geqslant 0\}$具有独立增量,

(iii) $P\{N(t+h)-N(t)\geqslant 2\}=o(h)$,

(iv) $P\{N(t+h)-N(t)=1\}=\lambda(t)h+o(h)$。

若令

$$m(t) = \int_0^t \lambda(s)ds$$

则可证明

(2.4.1) $P\{N(t+s)-N(t)=n\}$

$$=\exp\{-(m(t+s)-m(t))\}[m(t+s)-m(t)]^n/n!,\quad n\geqslant 0$$

即 $N(t+s)-N(t)$有均值为 $m(t+s)-m(t)$的泊松分布。

沿着定理 2.1.1 的证明路线，稍加修改可得(2.4.1)的证明：固定 t 且定义

$$P_n(s)=P\{N(t+s)-N(t)=n\}$$

则有

$P_0(s+h)=P\{N(t+s+h)-N(t)=0\}$

$=P\{$在$(t,t+s]$中没有事件，在$(t+s,t+s+h]$中没有事件$\}$

$=P\{$在$(t,t+s]$中没有事件$\}P\{$在$(t+s,t+s+h]$中没有事件$\}$

$=P_0(s)[1-\lambda(t+s)h+o(h)]$

其中最后第二个等式由公理(ii)得到，而最后的等式由公理(iii)与(iv)得到。因此，

$$\frac{P_0(s+h)-P_0(s)}{h}=-\lambda(t+s)P_0(s)+\frac{o(h)}{h}$$

令 $h\to 0$ 得

$$P_0'(s)=-\lambda(t+s)P_0(s)$$

或

$$\log P_0(s)=-\int_0^s\lambda(t+u)\,du$$

或

$$P_0(s)=e^{-[m(t+s)-m(t)]}$$

类似地可证(2.4.1)的其余结论，将它们留作一个习题。*

非齐次泊松过程的重要性在于不再要求平稳增量性，从而允许事件在某些时刻发生的可能性较之另一些时刻来得大。

当强度函数 $\lambda(t)$有界时，可以将非齐次泊松过程看作为一个齐次泊松过程的随机取样。具体地说，设 λ 满足

$$\lambda(t)\leqslant\lambda,\qquad 对一切\ t\geqslant 0$$

且考虑一个强度为 λ 的泊松过程。设此过程在时刻 t 发生的事件

---

* 事实上，定义 2.4.1 中的四个条件尚不充足，应补充假设 $\lambda(t)$ 连续及 $\lim_{h\to 0}P(N(t+h)-N(t)=0)=1$。——译者注

以概率 $\lambda(t)/\lambda$ 被计数，则被计数的事件构成的过程是有强度函数 $\lambda(t)$ 的非齐次泊松过程。容易从定义 2.4.1 得到此结论。例如，(i)，(ii)，(iii) 成立，因为对齐次的泊松过程它们也是成立的。公理(iv)成立，因为

$$\begin{aligned}&P\{\text{在}(t,t+h]\text{中有一个被计数的事件}\}\\&=P\{\text{在}(t,t+h]\text{中有一个事件}\}\frac{\lambda(t)}{\lambda}+o(h)\\&=\lambda h\frac{\lambda(t)}{\lambda}+o(h)\\&=\lambda(t)h+o(h)\end{aligned}$$

把非齐次泊松过程作为齐次泊松过程的取样的解释也给我们提供了理解命题 2.3.2 的另一条途径(或等价地，提供了证明命题 2.3.2 中的 $N(t)$ 服从泊松分布的另一条途径)。

例 2.4(a)　记录值。设 X_1，X_2，…是一列独立同分布的非负连续随机变量，其风险率函数为 $\lambda(t)$（即 $\lambda(t)=\frac{f(t)}{\bar{F}(t)}$，其中 f 与 F 分别是 X 的密度及分布函数）。若 $X_n>\max(X_1,\cdots,X_{n-1})$，其中 $X_0\equiv0$，我们说在时刻 n 产生了一个记录。若时刻 n 产生一个记录，则称 X_n 是一个记录值。以 $N(t)$ 记小于或等于 t 的记录值的个数，所以 $N(t)$ 是一计数过程，其中若 x 是一个记录值则说在时刻 x 一个事件发生。

我们说 $\{N(t),t\geqslant0\}$ 是一个非齐次的泊松过程，有强度函数 $\lambda(t)$。为验证这一结论，注意到 t 与 $t+h$ 之间有一个记录值当且仅当值超过 t 的第一个 X_i 必介于 t 与 $t+h$ 之间。但由风险率函数的定义，我们有（对 X_i 是哪一个变量取条件，比如说 $i=n$）：

$$P\{X_n\in(t,t+h)\mid X_n>t\}=\lambda(t)h+o(h)$$

由此证得断言。

例 2.4(b)　无穷个服务员的泊松排队系统（$M/G/\infty$）的输出过程。

实际上 $M/G/\infty$ 系统（即有无穷个服务员，泊松来到及共同的服务时间分布 G 的系统）的输出过程是强度函数为 $\lambda(t)=\lambda G(t)$ 的非齐次泊松过程。为证此结论，先证

(1) 在 $(s,s+t)$ 中离开的人数服从均值为 $\lambda\int_s^{s+t}G(y)dy$ 的泊松分布，且

(2) 在不相交的时间区间中离开的人数是独立的。

为证(1)，称在区间$(s,s+t)$内离开的顾客为I-型的，则在时刻y来到的顾客属于I-型的概率是

$$P(y)=\begin{cases}G(s+t-y)-G(s-y), & \text{当 } y<s\\ G(s+t-y), & \text{当 } s<y<s+t\\ 0, & \text{当 } y>s+t\end{cases}$$

因此，从命题2.3.2知这类离去的人数服从泊松分布，均值为

$$\lambda\int_0^{\infty}P(y)dy=\lambda\int_0^{s}(G(s+t-y)-G(s-y))dy+\lambda\int_s^{s+t}G(s+t-y)dy$$
$$=\lambda\int_s^{s+t}G(y)dy$$

为证明(2)，设I_1与I_2为不相交的时间区间，称在I_1内离开者为一个I-型顾客，称在I_2内离开者为Ⅱ-型顾客，其余的称为Ⅲ-型的。又从命题2.3.2或更确切地说是以它的有三种类型顾客的推广得知，在I_1与I_2中离去的人数(即I-型与Ⅱ-型顾客的人数)是独立的泊松随机变量。

利用(1)与(2)，验证输出过程满足非齐次泊松过程的全部公理化要求是件简单的事(它很像泊松过程的定义2.1.1蕴含定义2.1.2的证明)。

因为$t\to\infty$时，$\lambda(t)\to\lambda$，所以有趣的是注意到，长时间t(当$t\to\infty$)之后，极限输出过程是强度为λ的泊松过程。

2.5 复合泊松过程

称随机过程$\{X(t),t\geqslant 0\}$为复合泊松过程，若对$t\geqslant 0$它可以表示为：

$$X(t)=\sum_{i=1}^{N(t)}Y_i$$

其中$\{N(t),t\geqslant 0\}$是一泊松过程，$\{Y_i, i=1, 2, \cdots\}$是一族独立同分布的随机变量，且过程$\{N(t),t\geqslant 0\}$与$\{Y_i,i\geqslant 1\}$假定是独立的。

下面是复合泊松过程的一个例子。假定顾客按参数为λ的泊松过程进入一个商店，又假设各顾客所花的钱数形成一族独立同分布的随机变量。那么，以$X(t)$记到时间t顾客们在此商店所花费的总值，可见$\{X(t),t\geqslant 0\}$是一个复合泊松过程。

现在计算$X(t)$的矩母函数。

$$\phi_t(u)=E[\exp\{uX(t)\}]$$

$$=\sum_{n=0}^{\infty}E[\exp\{uX(t)\}|N(t)=n]P\{N(t)=n\}$$

$$=\sum_{n=0}^{\infty}E[\exp\{u(Y_1+\cdots+Y_n)\}|N(t)=n]e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

(2.5.1) $$=\sum_{n=0}^{\infty}E[\exp\{u(Y_1+\cdots+Y_n)\}]e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

(2.5.2) $$=\sum_{n=0}^{\infty}(E[\exp\{uY_1\}])^n e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

其中(2.5.1)得自$\{Y_1,Y_2,\cdots\}$与$N(t)$的独立性，而(2.5.2)则来自诸Y_i的独立性。因此，令

$$\phi_y(u)=E[e^{uY}]$$

从(2.5.2)我们得

$$\phi_t(u)=\sum_{n=0}^{\infty}\phi_y^n(u)e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

$$=\exp\{\lambda t(\phi_y(u)-1)\}$$

对上式求导易得

(2.5.3) $$E[X(t)]=\lambda tE[Y]$$

及

(2.5.4) $$\mathrm{Var}[X(t)]=\lambda tE[Y^2]$$

2.6 条件泊松过程

设Λ是具有分布G的正值随机变量，$\{N(t),t\geq0\}$是一计数过程，使得在已知$\Lambda=\lambda$的条件下$\{N(t),t\geq0\}$是参数为λ的泊松过程。于是，例如有

$$P\{N(t+s)-N(s)=n\}=\int_0^{\infty}e^{-\lambda t}\frac{(\lambda t)^n}{n!}dG(\lambda)$$

过程$\{N(t),t\geq0\}$称为条件泊松过程，因为在事件$\Lambda=\lambda$的条件下，它是参数为λ的泊松过程。然而，应该指出，$\{N(t),t\geq0\}$不

是一个泊松过程。例如，虽然它有平稳增量，但不具有独立增量。（为何不?）

让我们计算在已知 $N(t)=n$ 的条件下 Λ 的条件分布。对很小的 $d\lambda$

$$
\begin{aligned}
&P\{\Lambda\in(\lambda,\lambda+d\lambda)\mid N(t)=n\}\\
&=\frac{P\{N(t)=n\mid \Lambda\in(\lambda,\lambda+d\lambda)\}P\{\Lambda\in(\lambda,\lambda+d\lambda)\}}{P\{N(t)=n\}}\\
&=\frac{e^{-\lambda t}\dfrac{(\lambda t)^n}{n!}dG(\lambda)}{\displaystyle\int_0^\infty e^{-\lambda t}\frac{(\lambda t)^n}{n!}dG(\lambda)}
\end{aligned}
$$

从而在已知 $N(t)=n$ 的条件下，Λ 的条件分布为

$$
P\{\Lambda\leqslant x\mid N(t)=n\}=\frac{\displaystyle\int_0^x e^{-\lambda t}(\lambda t)^n dG(\lambda)}{\displaystyle\int_0^\infty e^{-\lambda t}(\lambda t)^n dG(\lambda)}
$$

例 2.6(a) 假设在某地区某个季节中地震发生的平均强度不是 λ_1 就是 λ_2，这取决于一些暂不清楚的因素。还假设百分之 $100p$ 的年份强度是 λ_1，而其余年份则是 λ_2，这种情况的一种简单模型是假设 $\{N(t),0\leqslant t<\infty\}$ 是一条件泊松过程，其中 Λ 分别以概率 p 与 $1-p$ 等于 λ_1 与 λ_2。已知在一个季节起先的 t 时间内发生 n 次地震，则这是一个 λ_1 季节的概率为

$$
P\{\Lambda=\lambda_1\mid N(t)=n\}=\frac{pe^{-\lambda_1 t}(\lambda_1 t)^n}{pe^{-\lambda_1 t}(\lambda_1 t)^n+e^{-\lambda_2 t}(\lambda_2 t)^n(1-p)}
$$

再对 $\Lambda=\lambda_1$ 或 $\Lambda=\lambda_2$ 取条件，可见在已知 $N(t)=n$ 的条件下，从 t 到下次地震的时间有分布

$$
\begin{aligned}
&P\{\text{从 } t \text{ 到下次地震的时间}\leqslant x\mid N(t)=n\}\\
&=\frac{p(1-e^{-\lambda_1 x})e^{-\lambda_1 t}(\lambda_1 t)^n+(1-e^{-\lambda_2 x})e^{-\lambda_2 t}(\lambda_2 t)^n(1-p)}{pe^{-\lambda_1 t}(\lambda_1 t)^n+e^{-\lambda_2 t}(\lambda_2 t)^n(1-p)}
\end{aligned}
$$

习 题

1. 证明泊松过程的定义 2.1.1 蕴含定义 2.1.2。
2. 证明定义 2.1.2 蕴含定义 2.1.1 的又一个方法：

(a) 利用定义 2.1.2，证明

$$P_0(t+s)=P_0(t)P_0(s)$$

(b) 用(a)推出来到间隔 X_1，X_2，…是具有参数 λ 的独立指数随机变量。

(c) 用(b)证明 $N(t)$ 服从均值为 λt 的泊松分布。

3. 对泊松过程证明：当 $s<t$ 时

$$P\{N(s)=k|N(t)=n\}=\binom{n}{k}\left(\frac{s}{t}\right)^k\left(1-\frac{s}{t}\right)^{n-k},$$

$$k=0,\ 1,\ \cdots,\ n$$

4. 设 $\{N(t),t\geqslant 0\}$ 是参数为 λ 的泊松过程。计算 $E[N(t)\cdot N(t+s)]$。

5. 假设 $\{N_1(t),t\geqslant 0\}$ 与 $\{N_2(t),t\geqslant 0\}$ 分别是参数为 λ_1 与 λ_2 的独立泊松过程。证明 $\{N_1(t)+N_2(t),t\geqslant 0\}$ 是参数为 $\lambda_1+\lambda_2$ 的泊松过程，且证明此过程的第一个事件来自 $\{N_1(t),t\geqslant 0\}$ 的概率是 $\lambda_1/(\lambda_1+\lambda_2)$，且与此事件发生的时间独立。

6. 一部机器运转需要两类部件。我们已备有 n 个 1-型的部件和 m 个 2-型的部件。i-型部件在失效之前持续运转的时间是参数为 μ_i 的指数变量。若部件失效时即以备件中取一同类部件更换，求机器正常运转时间的平均长度，即求 $E\left[\min\left(\sum_{i=1}^{n}X_i,\sum_{i=1}^{m}Y_i\right)\right]$，其中 X_i (Y_i) 是参数为 μ_1 (μ_2) 的指数变量。

7. 计算 S_1，S_2，S_3 的联合分布。

8. 产生一个泊松随机变量。设 U_1，U_2，…是独立的 $(0,1)$ 上随机变量的均匀分布。

(a) 若 $X_i=(-\log U_i)/\lambda$，证明 X_i 服从参数为 λ 的指数分布。

(b) 若 N 定义为满足下式之 n 值：

$$\prod_{i=1}^{n}U_i\geqslant e^{-\lambda}>\prod_{i=1}^{n+1}U_i$$

其中 $\prod_{i=1}^{0}U_i\equiv 1$，利用(a)证明 N 服从均值为 λ 的泊松分布。试与第一章习题 12 比较。

9. 考虑一个从底层起动上升的电梯。以 N_i 记在第 i 层进入电梯的人数。假定 N_i 相互独立，且 N_i 是均值为 λ_i 的泊松变量。在第 i 层进入的各个人相互独立地以概率 P_{ij} 在 j 层离开电梯，$\sum_{j>i}P_{ij}=1$。令 $O_j=$ 在第 j 层离开电梯的人数。

(a) 计算 $E[O_j]$。

(b) O_j的分布是什么?

(c) O_j与O_k的联合分布是什么?

10. 考虑一个具有r个面的骰子，且假定每掷一次只呈现其一面，第i面以概率P_i出现，$\sum_{i=1}^{r} P_i = 1$。给定数n_1，$\cdots n_r$，以N_i记直到第i面出现n_i次时所需掷次数，$i=1$，2，$\cdots$，r。令

$$N = \min_{i=1,\cdots,r} N_i$$

于是N是对某个$i=1$，2，$\cdots$，r，第i面出现n_i次时所需掷次数。

(a) N_i的分布是什么?

(b) 诸N_i独立吗?

今假设抛掷是在参数为$\lambda=1$的泊松过程所产生的随机来到时刻进行的。以T_i记直到第i面出现n_i次时所需时间，$i=1$，2，$\cdots$，r。令

$$T = \min_{i=1,\cdots,r} T_i$$

(c) T_i的分布是什么?

(d) 诸T_i独立吗?

(e) 导出$E[T]$的表达式。

(f) 用(e)导出$E[N]$的表达式。

11. 设要做的试验次数是一个均值为λ的泊松随机变量。每次试验有n个可能结果，出现第i个结果的概率为P_i，$\sum_{i=1}^{n} P_i = 1$。各次试验相互独立以X_j记恰发生j次的结果的个数，$j=0$，1，$\cdots$。计算$E[X_j]$，$\mathrm{Var}(X_j)$。

12. 设$X_1, X_2, \cdots, X_n$是独立的有共同密度函数f的连续随机变量。以$X_{(i)}$记X_1，$\cdots$，X_n的第i个最小者。

(a) 注意到为使$X_{(i)}$恰好等于x，诸X中的$i-1$个须小于x，有一个必须等于x，而其它的$n-i$个须都大于x。利用这个事实来论证$f_{X_{(i)}}$的密度为：

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!}(F(x))^{i-1}(\overline{F}(x))^{n-i}f(x)$$

(b) $X_{(i)}$小于x当且仅当有多少个X小于x?

(c) 利用(b)求$P\{X_{(i)} \leqslant x\}$的表达式。

(d) 利用(a)与(c)建立恒等式：对$0 \leqslant y \leqslant 1$，

$$\sum_{k=i}^{n} \binom{n}{k} y^k (1-y)^{n-k} = \int_0^y \frac{n!}{(i-1)!(n-i)!} x^{i-1}(1-x)^{n-i} dx$$

(e) 以S_i记泊松过程$\{N(t), t \geqslant 0\}$的第i个事件发生的时刻。求

$$E[S_i | N(t) = n] = \begin{cases} & , i \leqslant n \\ & , i > n \end{cases}$$

13. 以 $U_{(1)}$，…，$U_{(n)}$记 n 个独立的(0,1)上的均匀分布的随机变量的顺序统计量。证明：在已知$U_{(n)}=y$ 的条件下，$U_{(1)}$，…，$U_{(n-1)}$的分布与（$n-1$）个独立的（0，y）上均匀分布的随机变量的顺序统计量的分布相同。

14. 载有顾客的客车按参数为 λ 的泊松过程来到一有无穷多个服务员的排队系统。以 G 记服务时间分布。一辆客车载有 j 个顾客的概率为 α_j，$j=1$，2，…。以 $X(t)$记到 t 时已被服务过的顾客人数。

(a) $E[X(t)]=?$

(b) $X(t)$服从泊松分布吗?

15. 假设泊松过程的每个事件被分成类型 1，2，…，k。若事件在时刻 s 发生，则它与其它事件相互独立以概率 $P_i(s)$ 划归为 i 型，$i=1$，2，…，k，$\sum_{i=1}^{k} P_i(s) = 1$ 。以 $N_i(t)$ 记在 $[0,t]$ 内 i 型事件来到的个数。证明：$N_i(t)$，$i=1$，2，…，k，相互独立，且 $N_i(t)$ 服从均值为 $\lambda\int_0^t P_i(s)ds$ 的泊松分布，其中 λ 是泊松过程的参数。

16. 个体按照参数为 λ 的泊松过程进入系统。每个来到的个体相互独立地经历系统的状态。以 $\alpha_i(s)$ 记一个体来到后经过时间 s 处于状态 i 的概率。以 $N_i(t)$ 记在时刻 t 处于状态 i 的个体数。证明：$N_i(t)$，$i\geqslant 1$，相互独立且 $N_i(t)$ 是泊松变量，均值等于

λE[一个体从进入系统到 t 时止处于状态 i 的时间]。

17. 假设汽车按参数为 λ 的泊松过程进入一条单向行驶的无限长的高速公路，进入的第 i 辆车以速度 V_i 行驶。假定诸 V_i 是独立的正随机变量，有共同分布 F。推导出在时刻 t 位于区间(a,b)内的汽车数的分布，假定一辆车超过另一辆车时，不占用任何时间。

18. 计算在已知 $S_n=t$ 的条件下 S_1，…，S_n 的条件分布。

19. 计算例 2.3(c) 中 $D(t)$ 的矩母函数。

20. 证明引理 2.3.3。

21. 对非齐次泊松过程完成 $N(t+s)-N(t)$ 是均值为 $m(t+s)-m(s)$ 的泊松变量的证明。

22. 以 T_1，T_2，…记强度函数为 $\lambda(t)$的非齐次泊松过程的事件的来到间隔。

(a) 诸T_i 独立否?

(b) 诸T_i 同分布吗?

(c) 求T_1的分布；

(d) 求T_2的分布。

23. 考虑一非齐次泊松过程 $\{N(t),t\geqslant 0\}$，对一切t，其$\lambda(t)>0$。令

$$N^*(t)=N(m^{-1}(t))$$

证明：$\{N^*(t),t\geqslant 0\}$ 是强度$\lambda=1$的泊松过程。

24. (a) 设 $\{N(t),t\geqslant 0\}$ 是一均值函数为$m(t)$的非齐次泊松过程，在已知$N(t)=n$的条件下证明来到时刻不讲次序与n个独立同分布函数

$$F(x)=\begin{cases}\dfrac{m(x)}{n(t)}, & x\leqslant t\\ 1, & x>t\end{cases}$$

的随机变量有相同分布。

(b) 假设工人按照均值函数为$m(t)$的非齐次泊松过程遭到事故。又假设每个受伤者停工时间是分布为F的随机变量。记$X(t)$为时刻t停工的工人数。计算$E[X(t)]$及$\mathrm{Var}(X(t))$。

25. 一个二维泊松过程是在平面上随机发生的事件构成的过程，使得：(i) 对于任一面积为A的区域，其中的事件数服从均值为λA的泊松分布；(ii) 不相交的区域中的事件数是相互独立的。对此过程，考虑平面上任意一点，以X记它与最相近的事件间的距离（这里的距离即通常的欧几里德距离）。证明：

(a) $P\{X>t\}=e^{-\lambda\pi t^2}$

(b) $E[X]=1/2\sqrt{\lambda}$

26. 对复合泊松过程计算$\mathrm{Cov}(X(s),X(t))$。

27. 验证(2.5.4)式与(2.5.3)式。

28. 对一个条件泊松过程：

(a) 解释为什么条件泊松过程有平稳增量但无独立增量。

(b) 在已知 $\{N(s),0\leqslant s\leqslant t\}$，即过程直到$t$时的历史资料条件下，计算$\Lambda$的条件分布，并且证明它只依赖于$N(t)$。解释为什么会是这样的。

(c) 在已知$N(t)=n$的条件下，计算t之后第一个事件发生的时刻的条件分布。

(d) 计算

$$\lim_{h\to 0}\frac{P\{N(h)\geqslant 1\}}{h}$$

(e) 以X_1，X_2，…记来到间隔。它们独立吗？它们同分布吗？

29. 考虑一个条件泊松过程，其中 Λ 服从参数为 m 与 α 的 Γ-分布，即密度为

$$g(\lambda)=\alpha e^{-\lambda\alpha}(\lambda\alpha)^{m-1}/(m-1)!,\quad 0<\lambda<\infty$$

(a) 证明：

$$P\{N(t)=n\}=\binom{m+n-1}{n}\left(\frac{\alpha}{\alpha+t}\right)^m\left(\frac{t}{\alpha+t}\right)^n,\quad n\geqslant 0$$

(b) 证明：在已知 $N_1(t)=n$ 的条件下，Λ 的条件分布仍是 Γ-分布，参数为 $m+n$，$\alpha+t$。

(c) $\lim_{h\to 0}P\{N(t+h)-N(t)=1|N(t)=n\}/h$ 是什么？

参考文献

在许多教科书中都有泊松过程的论述，例如：

1. E. Cinlar, *Introduction to Stochastic Processes*, Prentice Hall, Englewood Cliffs, N.J., 1975.
2. E. Parzen, *Stochastic Processes*, Holden-Day, San Francisco, 1962。（帕尔逊，随机过程，邓永录、杨振明译，高等教育出版社，1987。）
3. S. Ross, *Introduction to Probability Models*, 2nd ed., Academic Press, New York, 1981.

3
更新理论

3.1 引言与基本定义

上一章我们看到泊松过程的来到间隔是独立同分布的指数随机变量。一种自然的推广是考虑来到间隔独立同分布，但分布函数任意的计数过程。这样的计数过程称为更新过程。

正式地，设$\{X_n,n=1,2,\cdots\}$是一列非负的随机变量，具有共同的分布 F，且为了避免显而易见的平凡情形，假设 $F(0)=P\{X_n=0\}<1$。我们将把 X_n 解释为第 $n-1$ 个与第 n 个事件之间的时间。以

$$\mu=E[X_n]=\int_0^\infty xdF(x)$$

记相继发生的两事件的间隔之均值，且注意到从假定 $X_n\geqslant 0$ 与 $F(0)<1$，可得 $0<\mu\leqslant\infty$。令

$$S_0=0,\quad S_n=\sum_{i=1}^n X_i,\quad n\geqslant 1$$

可知 S_n 是第 n 个事件发生的时刻。因为到时刻 t 已发生的事件个数等于使第 n 个事件在时间 t 或 t 之前发生的最大的 n 的值，所以到时刻 t 已发生的事件的个数 $N(t)$ 为

(3.1.1) $$N(t)=\sup\{n: S_n\leqslant t\}$$

定义 3.1.1 计数过程 $\{N(t), t \geqslant 0\}$ 称为更新过程。

我们将通用事件与更新这两个词，因而称第 n 次更新在时刻 S_n 发生。由于间隔是独立同分布的，所以在各个更新时刻此过程在概率意义上重新开始。

我们将要回答的第一个问题是在有限时间内是否会有无限多次更新发生。为了证明这是不可能的，我们注意到，由强大数定律可知，以概率 1，当 $n \to \infty$ 时

$$\frac{S_n}{n} \to \mu$$

但因 $\mu > 0$，所以这意味着当 n 趋于无穷时，S_n 必将趋于无穷。于是至多只有有限多个 n 能使 S_n 小于或等于 t，因此，由(3.1.1)知 $N(t)$ 必是有限的，所以可写成

$$N(t) = \max\{n : S_n \leqslant t\}$$

3.2 $N(t)$ 的分布

$N(t)$ 的分布至少在理论上能得到，只要先注意到这么一个重要的关系：到时刻 t 为止的更新次数大于或等于 n 当且仅当在 t 之前或在时刻 t 发生第 n 次更新，即

(3.2.1) $$N(t) \geqslant n \quad \Leftrightarrow \quad S_n \leqslant t$$

从 (3.2.1) 得到

(3.2.2) $$\begin{aligned} P\{N(t)=n\} &= P\{N(t) \geqslant n\} - P\{N(t) \geqslant n+1\} \\ &= P\{S_n \leqslant t\} - P\{S_{n+1} \leqslant t\} \end{aligned}$$

既然随机变量 X_i，$i \geqslant 1$，是独立的且有共同的分布函数 F，所以 $S_n = \sum_{i=1}^{n} X_i$ 的分布 F_n 为 F 自身的 n 次卷积。因而，从 (3.2.2) 我们得到

$$P\{N(t) = n\} = F_n(t) - F_{n+1}(t)$$

令
$$m(t) = E[N(t)]$$

称 $m(t)$ 为更新函数，更新理论的大部分内容涉及到确定 $m(t)$ 的性质。$m(t)$ 与 F 之间的关系由下列命题给出。

命题 3.2.1

$$m(t)=\sum_{n=1}^{\infty}F_n(t) \tag{3.2.3}$$

证明

$$N(t)=\sum_{n=1}^{\infty}I_n$$

其中

$$I_n=\begin{cases}1, & \text{若第 } n \text{ 次更新发生在}[0,t]\text{内} \\ 0, & \text{其它}\end{cases}$$

因此，

$$\begin{aligned}E[N(t)]&=E\Big[\sum_{n=1}^{\infty}I_n\Big]\\&=\sum_{n=1}^{\infty}E[I_n]\\&=\sum_{n=1}^{\infty}P\{I_n=1\}\\&=\sum_{n=1}^{\infty}P\{S_n\leqslant t\}\\&=\sum_{n=1}^{\infty}F_n(t)\end{aligned}$$

其中由于 I_n 非负，求期望与求和顺序交换是合理的。

下一个命题证明 $N(t)$ 有有限的期望值。

命题 3.2.2

$$\text{对一切 } 0\leqslant t\leqslant\infty,\quad m(t)<\infty$$

证明 因 $P\{X_n=0\}<1$，由概率的连续性可知存在一个$\alpha>0$，使得 $P\{X_n\geqslant\alpha\}>0$。现在定义一个关连的更新过程$\{\overline{X}_n,n\geqslant1\}$如下：

$$\overline{X}_n=\begin{cases}0, & \text{若 } X_n<\alpha \\ \alpha, & \text{若 } X_n\geqslant\alpha\end{cases}$$

且令 $\overline{N}(t)=\sup\{n:\overline{X}+\cdots+\overline{X}_n\leqslant t\}$。易知，对此过程更新只能在时刻 $t=n\alpha$ 发生，$n=0,1,2,\cdots$，且在这些时刻的更新次数是独立的几何随机变

量，其均值为

$$\frac{1}{P\{X_n \geqslant \alpha\}}$$

于是

$$E[\overline{N}(t)] \leqslant \frac{(t/\alpha + 1)}{P\{X_n \geqslant \alpha\}} < \infty$$

且因 $\overline{X}_n \leqslant X_n$ 蕴含着 $\overline{N}(t) \geqslant N(t)$，所以结论得证。

注记　以上的证明还表明了对一切 $t \geqslant 0$，$r > 0$，$E[N^2(t)] < \infty$。

3.3 若干极限定理

若以 $N(\infty) = \lim\limits_{t \to \infty} N(t)$ 记所发生的更新总数，容易看到以概率 1，

$$N(\infty) = \infty$$

这是因为使所发生的更新总数 $N(\infty)$ 为有限的唯一方法是有一个来到间隔为无穷大。所以

$$\begin{aligned} P\{N(\infty) < \infty\} &= P\{X_n = \infty, \text{对某个 } n\} \\ &= P\{\bigcup_{n=1}^{\infty} \{X_n = \infty\}\} \\ &\leqslant \sum_{n=1}^{\infty} P\{X_n = \infty\} \\ &= 0 \end{aligned}$$

于是当 t 趋于无穷时 $N(t)$ 趋于无穷。然而，有兴趣的是了解 $N(t)$ 趋于无穷的速度，即我们要能说出一些关于 $\lim\limits_{t \to \infty} \frac{N(t)}{t}$ 的情况。

作为确定 $N(t)$ 增长速率的准备，让我们先考虑随机变量 $S_{N(t)}$。用话来说，这随机变量表示什么呢？我们以具体情形作归纳，例如，假设 $N(t) = 3$，则 $S_{N(t)} = S_3$ 表示第三个事件发生的时刻。由于到时刻 t 为止已发生的事件只有三个，所以 S_3 也表示在

t 之前(或在时刻 t)最后一个事件发生的时刻,事实上这就是 $S_{N(t)}$ 所表示的:在时刻 t 之前或在时刻 t 最后一次更新的时刻。类似的推理可知 $S_{N(t)+1}$ 表示时刻 t 之后第一次更新的时刻(见图 3.3.1)。

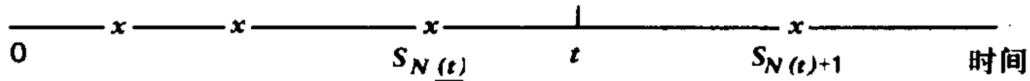


图 3.3.1

现在我们可以证明下列

命题 3.3.1　以概率 1,当 $t\to\infty$ 时

$$\frac{N(t)}{t}\to\frac{1}{\mu}$$

证明　因 $S_{N(t)}\leqslant t<S_{N(t)+1}$,可见

$$\frac{S_{N(t)}}{N(t)}\leqslant\frac{t}{N(t)}<\frac{S_{N(t)+1}}{N(t)} \tag{3.3.1}$$

然而,因 $S_{N(t)}/N(t)$ 是前 $N(t)$ 个来到间隔的平均值,由强大数定律得出,当 $N(t)\to\infty$ 时,$S_{N(t)}/N(t)\to\mu$。但由于 $t\to\infty$ 时 $N(t)\to\infty$,所以 $t\to\infty$ 时

$$\frac{S_{N(t)}}{N(t)}\to\mu$$

又写

$$\frac{S_{N(t)+1}}{N(t)}=\left[\frac{S_{N(t)+1}}{N(t)+1}\right]\left[\frac{N(t)+1}{N(t)}\right]$$

由同样的推理得 $t\to\infty$ 时

$$\frac{S_{N(t)+1}}{N(t)}\to\mu$$

现在由(3.3.1)知结论成立,因为 $t/N(t)$ 介于两变量之间,它们在 $t\to\infty$ 时都收敛于 μ。

命题 3.3.1 说以概率 1,长时间后更新发生的速率将等于 $1/\mu$。为此称 $1/\mu$ 为更新过程的速率。

我们要证明更新的平均速度的期望 $m(t)/t$ 也收敛于 $1/\mu$。然而在给出证明之前,我们觉得先离开本题,讨论一下停时与瓦尔德等式是很有用的。

3.3.1　瓦尔德等式

设 X_1,X_2,…为一列独立随机变量。我们有以下定义。

定义 整值随机变量 N 称为序列 $X_1, X_2, \cdots$ 的停时，若对一切 $n=1, 2, \cdots$，事件 $\{N=n\}$ 与 $X_{n+1}, X_{n+2}, \cdots$ 独立。

直观上看，我们依次观察诸 X_n，以 N 记在停止观察之前所观察的次数。若 $N=n$，则在观察 $X_1, \cdots, X_n$ 之后与观察 $X_{n+1}, X_{n+2}, \cdots$ 之前我们停止观察。

例 3.3(a) 设 X_n, $n=1, 2, \cdots$，相互独立且使得

$$P\{X_n = 0\} = P\{X_n = 1\} = \frac{1}{2}, \qquad n = 1,2,\cdots$$

如果我们令

$$N=\min\{n: X_1+\cdots+X_n=10\}$$

则 N 是一个停时。我们可以将 N 看作连续地抛掷一枚均匀硬币的试验的停时，试验在正面出现次数达到 10 次时停止。

例 3.3(b) 设 X_n, $n=1, 2, \cdots$，相互独立且使得

$$P\{X_n = -1\} = P\{X_n = 1\} = \frac{1}{2}$$

则

$$N = \min\{n: X_1 + \cdots + X_n = 1\}$$

是一个停时。它可以看作为一个赌徒的停时，他在每局赌博中等可能地赢得或输掉一元，且决定一旦领先就停止。（下一章中将证明，以概率 1，N 是有限的。）

定理 3.3.2（瓦尔德等式）若 $X_1, X_2, \cdots$ 是独立同分布的随机变量，期望有限，且 N 是 $X_1, X_2, \cdots$ 的停时，使得 $E[N] < \infty$，则

$$E\Big[\sum_{n=1}^{N} X_n\Big] = E[N]E[X]$$

证明 令

$$I_n=\begin{cases}1, & 若\ N\geqslant n\\ 0, & 若\ N<n\end{cases}$$

我们有

$$\sum_{n=1}^{N} X_n = \sum_{n=1}^{\infty} X_n I_n$$

因此，

(3.3.2) $$E\left[\sum_{n=1}^{N}X_n\right]=E\left[\sum_{n=1}^{\infty}X_nI_n\right]=\left[\sum_{n=1}^{\infty}E\ [X_nI_n]\right]$$

然而 $I_n=1$ 当且仅当我们连续观察 X_1, X_2, …, X_{n-1}之后不停下来。所以 I_n 由 X_1, …, X_{n-1}决定而与 X_n 独立。于是从(3.3.2)我们得到

$$\begin{aligned}E\left[\sum_{n=1}^{N}X_n\right]&=\sum_{n=1}^{\infty}E[X_n]E[I_n]\\&=E[X]\sum_{n=1}^{\infty}E[I_n]\\&=E[X]\sum_{n=1}^{\infty}P\{N\geqslant n\}\\&=E[X]E[N]\end{aligned}$$

注记　在(3.3.2)式中，我们未说明理由就变换了期望与和号的顺序。为论证其可交换性，以 X_i 的绝对值替换 X_i，此时因各项都非负交换顺序是合理的。然而根据勒贝格控制收敛定理，这就蕴含了原来的交换顺序是允许的。

对于例 3.3(a) 来说，瓦尔德等式蕴含了

$$E[X_1+\cdots+X_N]=\frac{1}{2}EN$$

然而，由 N 的定义 $X_1+\cdots+X_N=10$，从而 $E[N]=20$。

将瓦尔德等式的结论应用到例 3.3 (b) 会有等式 $E[X_1+\cdots+X_N]=E[N]E[X]$。然而 $X_1+\cdots+X_N=1$ 及 $E[X]=0$，从而得出矛盾。所以瓦尔德等式不可应用，这就得出结论 $E[N]=\infty$。

3.3.2　回到更新理论

以 X_1, X_2, …记一更新过程的来到间隔，让我们在 t 之后的第一个更新即 $N(t)+1$ 次更新时刻停止。为了证实 $N(t)+1$ 是序列 X_t 的停时，注意到

$$\begin{aligned}N(t)+1=n&\Leftrightarrow N(t)=n-1\\&\Leftrightarrow X_1+\cdots+X_{n-1}\leqslant t,\quad X_1+\cdots+X_N>t\end{aligned}$$

于是事件 $\{N(t)+1=n\}$ 只依赖于 X_1, …, X_n，故与 X_{n+1}, …独立；因此 $N(t)+1$ 是一个停时。从瓦尔德等式我们得到，当 $E[X]<\infty$ 时

$$E[X_1+\cdots+X_{N(t)+1}]=E[X]E[N(t)+1]$$

或等价地有下列

系 3.3.3 若 $\mu<\infty$，则

(3.3.3) $$E[S_{N(t)+1}]=\mu[m(t)+1]$$

现在我们能证明下列

定理 3.3.4（基本更新定理）当 $t\to\infty$ 时

$$\frac{m(t)}{t}\to\frac{1}{\mu}\quad（其中\frac{1}{\infty}\equiv 0）$$

证明 首先假设 $\mu<\infty$。现在有(见图 3.3.1)

$$S_{N(t)+1}>t$$

取期望并用系 3.3.3 得

$$\mu[m(t)+1]>t$$

这蕴含了

(3.3.4) $$\liminf_{t\to\infty}\frac{m(t)}{t}\geqslant\frac{1}{\mu}$$

以另一方面，固定一常数 M，且定义一个新的更新过程 $\{\overline{X}_n, n=1, 2, \cdots\}$ 如下：

$$\overline{X}_n=\begin{cases}X_n, & 当 X_n\leqslant M, \quad n=1, 2, \cdots \\ M, & 当 X_n>M\end{cases}$$

令 $\overline{S}_n=\sum_{i=1}^{n}\overline{X}_i$ 及 $\overline{N}(t)=\sup\{n: \overline{S}_n\leqslant t\}$。由于这个截尾更新过程的来到间隔时间以 M 为界，我们得到

$$\overline{S}_{N(t)+1}\leqslant t+M$$

因此由系 3.3.3，

$$[\overline{m}(t)+1]\mu_M\leqslant t+M$$

其中 $\mu_M=E[\overline{X}_n]$，于是

$$\limsup_{t\to\infty}\frac{\overline{m}(t)}{t}\leqslant\frac{1}{\mu_M}$$

而因 $\overline{S}_n\leqslant S_n$，得 $\overline{N}(t)\geqslant N(t)$ 及 $\overline{m}(t)\geqslant m(t)$，于是

(3.3.5) $$\limsup_{t\to\infty}\frac{m\ (t)}{t}\leqslant\frac{1}{\mu_M}$$

令 $M\to\infty$ 得

(3.3.6) $$\limsup_{t\to\infty}\frac{m(t)}{t}\leqslant\frac{1}{\mu}$$

从(3.3.4)及(3.3.6)得结论成立。

当 $\mu=\infty$，我们再考虑截尾过程；由于当 $M\to\infty$ 时 $\mu_M\to\infty$，从(3.3.5)知结论成立。

注记　初看一下，仿佛这个基本更新定理应当是命题 3.3.1 的一个简单推论。也就是由于平均更新速率以概率 1 收敛于 $1/\mu$，这不就应蕴含了平均更新速率的期望也收敛于 $1/\mu$ 吗？然而我们必须小心；考虑以下例子。

设 U 是(0,1)上均匀分布的随机变量。定义随机变量 Y_n，$n\geqslant1$，如下：

$$Y_n=\begin{cases}0, & 若U>1/n\\ n, & 若U\leqslant1/n\end{cases}$$

因为以概率 1，U 大于 0，故得对一切充分大的 n，Y_n 将等于 0，即对一切大到使 $\frac{1}{n}<U$ 的 n，Y_n 将为 0。因此，以概率 1，$n\to\infty$ 时

$$Y_n\to0$$

然而，

$$E[Y_n]=nP\{U\leqslant\frac{1}{n}\}=n\cdot\frac{1}{n}=1$$

所以，即使随机变量序列 Y_n 收敛于 0，Y_n 的期望值却恒等于 1。

我们以证明 $t\to\infty$ 时 $N(t)$ 为渐近正态分布来结束本节。为了证明这一结果，我们不仅使用中心极限定理(以证明 S_n 是渐近正态的)且要使用关系式

(3.3.7) $$N(t)<n\quad\Leftrightarrow\quad S_n>t$$

定理 3.3.5　设来到间隔的均值 μ 和方差 σ^2 有限，则 $t\to\infty$ 时

$$P\left\{\frac{N(t)-t/\mu}{\sigma\sqrt{t/\mu^3}}<y\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}e^{-x^2/2}dx$$

证明　令 $r_t=t/\mu+y\sigma\sqrt{t/\mu^3}$，则

$$P\left\{\frac{N(t)-t/\mu}{\sigma\sqrt{t/\mu^3}}<y\right\}=P\{N(t)<r_t\}$$

$$
\begin{aligned}
&= P\{S_{r_t} > t\} \qquad \text{(由 3.3.7)} \\
&= P\left\{\frac{S_{r_t} - r_t\mu}{\sigma\sqrt{r_t}} > \frac{t - r_t\mu}{\sigma\sqrt{r_t}}\right\} \\
&= P\left\{\frac{S_{r_t} - r_t\mu}{\sigma\sqrt{r_t}} > -y\left(1 + \frac{y\sigma}{\sqrt{t\mu}}\right)^{-1/2}\right\}
\end{aligned}
$$

现在由中心极限定理，当 t（于是 r_t）趋于∞时，$(S_{r_t} - r_t\mu)/\sigma\sqrt{r_t}$ 收敛于均值为 0 方差为 1 的正态随机变量。又因 $t \to \infty$ 时

$$
-y\left(1 + \frac{y\sigma}{\sqrt{t\mu}}\right)^{-1/2} \to -y
$$

可见

$$
P\left\{\frac{N(t) - t/\mu}{\sigma\sqrt{t/\mu^3}} < y\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-y}^{\infty} e^{-x^2/2} dx
$$

又因

$$
\int_{-y}^{+\infty} e^{-x^2/2} dx = \int_{-\infty}^{y} e^{-x^2/2} dx
$$

故结论成立。

注记

(i) 在上面的论证中有点小困难，因为要使用关系式(3.3.7)，r_t 应是一个整数。然而使上面的论证严格并不太困难。

(ii) 定理 3.3.5 说，$N(t)$ 是渐近正态的，均值为 t/μ，方差为 $t\sigma^2/\mu^3$。

3.4 关键更新定理及其应用

非负随机变量 X 称为格点的，若存在 $d \geqslant 0$ 使得 $\sum_{n=0}^{\infty} P\{X = nd\} = 1$。即 X 是格点的，若 X 只取某个非负数 d 的整数倍，具有这性质的最大的 d 称为 X 的周期。若 X 是格点的，F 是 X 的分布函数，则我们称 F 是格点的。

我们不加证明地叙述下列定理。

定理 3.4.1（布莱克威尔定理）

(i) 若 F 不是格点的，则对一切 $a\geqslant 0$，$t\to\infty$ 时

$$m(t+a)-m(t)\to a/\mu$$

(ii) 若 F 是格点的，周期为 d，则 $n\to\infty$ 时

$$E[\text{在时刻 } nd \text{ 更新的次数}]\to d/\mu$$

布莱克威尔定理说，若 F 不是格点的，则在一远离原点的长为 a 的区间中更新次数的期望近似于 a/μ。这是十分直观的，因为远离原点时初始影响好像耗尽了，所以

$$g(a)\equiv\lim_{t\to\infty}[m(t+a)-m(t)] \tag{3.4.1}$$

应当存在。然而，如果上述极限确实存在，作为基本更新定理的一个简单推论它必等于 a/μ。为了明白这一点，首先注意

$$\begin{aligned}g(a+b)&=\lim_{t\to\infty}[m(t+a+b)-m(t)]\\&=\lim_{t\to\infty}[m(t+a+b)-m(t+a)\\&\qquad\qquad+m(t+a)-m(t)]\\&=g(b)+g(a)\end{aligned}$$

然而，$g(a+b)=g(a)+g(b)$ 的唯一(递增的)解是

$$g(a)=ca,\qquad a>0$$

其中 c 是某个常数。为了证明 $c=1/\mu$，定义

$$\begin{aligned}x_1&=m(1)-m(0)\\x_2&=m(2)-m(1)\\&\vdots\\x_n&=m(n)-m(n-1)\\&\vdots\end{aligned}$$

则

$$\lim_{n\to\infty}x_n=c$$

蕴含了

$$\lim_{n\to\infty}\frac{x_1+\cdots+x_n}{n}=c$$

或

$$\lim_{n\to\infty}\frac{m(n)}{n}=c$$

因此，由基本更新定理，$c=1/\mu$。

当 F 是格点的,具有周期 d,则(3.4.1)中的极限不可能存在。因为此时更新只能发生在 d 的整数倍的时刻，于是在一远离原点的区间中更新次数的期望显然本来就不依赖于区间之长，而是依赖于它包含多少个形式为 nd，$n\geqslant 0$，的点。于是在格点的情形有关的极限是在时刻 nd 更新次数的期望的极限，且当 $\lim\limits_{n\to\infty}E$[在时刻 nd 更新的次数] 存在时，由基本更新定理它又必须等于 d/μ。若来到间隔总是正的，则布莱克威尔定理的(ii)说，在格点的情形

$$\lim_{n\to\infty}P\{\text{在时刻 } nd \text{ 更新}\}=d/\mu$$

设 h 是定义在[0,∞]上的一个函数。对任意 $a>0$，以 $\underline{m}_n(a)$ 与 $\overline{m}_n(a)$ 分别记 $h(t)$ 在区间 $(n-1)a\leqslant t\leqslant na$ 上的上确界与下确界。若对一切 $a>0$，$\sum\limits_{n=1}^{\infty}\overline{m}_n(a)$ 与 $\sum\limits_{n=1}^{\infty}\underline{m}_n(a)$ 有限，且

$$\lim_{a\to 0}a\sum_{n=1}^{\infty}\overline{m}_n(a)=\lim_{a\to 0}a\sum_{n=1}^{\infty}\underline{m}_n(a)$$

称 h 为直接黎曼可积。h 为直接黎曼可积的一个充分条件是

(i) 对一切 $t\geqslant 0$，$h(t)\geqslant 0$，

(ii) $h(t)$ 非增，

(iii) $\int_0^{\infty}h(t)dt<\infty$。

我们将不加证明地叙述下列定理，它以关键更新定理著称。

定理 3.4.2（关键更新定理）若 F 不是格点的，且若 $h(t)$ 直接黎曼可积，则

$$\lim_{t\to\infty}\int_0^t h(t-x)dm(x)=\int_0^{+\infty}\frac{h(t)}{\mu}dt$$

其中

$$m(x)=\sum_{n=1}^{\infty}F_n(x) \text{ 而 } \mu=\int_0^{+\infty}\overline{F}(t)dt$$

为了对关键更新定理有个感性认识，我们从布莱克威尔定理着手说明如下：由布莱克威尔定理，我们有

$$\lim_{t\to\infty}\frac{m(t+a)-m(t)}{a}=\frac{1}{\mu}$$

因此

$$\lim_{a\to 0}\lim_{t\to\infty}\frac{m(t+a)-m(t)}{a}=\frac{1}{\mu}$$

假定我们能论证极限可交换，就得到

$$\lim_{t\to\infty}\frac{dm(t)}{dt}=\frac{1}{\mu}$$

关键更新定理正是上述结果的一种形式。

能够证明布莱克威尔定理与关键更新定理是等价的。习题 9 要求读者以关键更新定理推出布莱克威尔定理；而以阶梯函数逼近直接黎曼可积函数可证明逆命题。

在第八章的 8.3 节中在 F 连续且其失效率函数有界并大于一个正数时，给出了布莱克威尔定理的一个概率证明。

关键更新定理是一个很重要且有用的结果。当要计算在时刻 t 的某些概率或期望 $g(t)$ 的极限时，便要用到它。在应用它时，我们采用的技巧是先关于在 t 之前（或在时刻 t）的最后一个更新的时刻取条件而导出一个 $g(t)$ 的方程。如我们将见到的，这就产生一个如下形式的方程：

$$g(t)=h(t)+\int_o^t h(t-x)dm(t)$$

我们从一个引理开始，它给出在 t 之前（或在时刻 t）的最后一个更新时刻 $S_{N(t)}$ 的分布。

引理 3.4.3

$$P\{S_{N(t)}\leqslant s\}=\overline{F}(t)+\int_o^s\overline{F}(t-y)dm(y),\quad t\geqslant s\geqslant 0$$

证明

$$P\{S_{N(t)}\leqslant s\}=\sum_{n=0}^{\infty}P\{S_n\leqslant s,S_{n-1}>t\}$$

$$
\begin{aligned}
&= \overline{F}(t) + \sum_{n=1}^{\infty} P\{S_n \leqslant s, S_{n+1} > t\} \\
&= \overline{F}(t) + \sum_{n=1}^{\infty} \int_o^{+\infty} P\{S_n \leqslant s, \\
&\qquad S_{n+1} > t \mid S_n = y\} dF_n(y) | \\
&= \overline{F}(t) + \sum_{n=1}^{\infty} \int_o^{s} \overline{F}(t-y) dF_n(y) \\
&= \overline{F}(t) + \int_o^{s} \overline{F}(t-y) d\Big(\sum_{n=1}^{\infty} F_n(y)\Big) \\
&= \overline{F}(t) + \int_o^{s} \overline{F}(t-y) dm(y)
\end{aligned}
$$

因为所有的项非负，所以积分与求和可换序。

注记

（1） 从引理 3.4.3 得

$P\{S_{N(t)} = 0\} = \overline{F}(t)$

$dF_{S_{N(t)}}(y) = \overline{F}(t-y)dm(y), \quad 0 < y < \infty$

（2） 为了直观地理解上式，假设 F 连续，具有密度 f，则

$m(y) = \sum_{n=1}^{\infty} F_n(y)$，从而，对 $y>0$

$$
\begin{aligned}
dm(y) &= \sum_{n=1}^{\infty} f_n(y)dy \\
&= \sum_{n=1}^{\infty} P\{\text{在}(y, y+dy)\text{ 中发生第 } n \text{ 次更新}\} \\
&= P\{\text{在}(y, y+dy)\text{ 中发生更新}\}
\end{aligned}
$$

所以，$S_{N(t)}$的概率密度为

$$
f_{S_{N(t)}}(y)dy = P\{\text{在}(y, y+dy)\text{中发生更新，且下一个来到间隔} > t-y\} = dm(y)\overline{F}(t-y)
$$

现在我们给出一些应用关键更新定理的例子。我们一再使用的技巧是关于 $S_{N(t)}$ 取条件。

3.4.1 交错更新过程

考虑一个系统，它有两个状态：开或关。最初它是开的且持续开的时间是 Z_1；而后关闭且持续闭的时间为 Y_1；之后又打开，时间为 Z_2；又关闭，时间为 Y_2；再打开等等。

我们假设随机向量 (Z_n, Y_n)，$n\geqslant 1$，独立同分布。因此，随机变量序列 $\{Z_n\}$ 与 $\{Y_n\}$ 都是独立同分布的；但允许 Z_n 与 Y_n 是相依的。换言之，每当过程打开时，一切就重新开始，但当它关闭时允许关闭的时间依赖于前一段打开的时间。

设 H 是 Z_n 的分布，G 是 Y_n 的分布，而 F 是 Z_n+Y_n，$n\geqslant 1$，的分布。又令

$$P(t)=P\{\text{时刻}\,t\,\text{系统开着}\}$$

定理 3.4.4 若 $E[Z_n+Y_n]<\infty$，且 F 不是格点的，则

$$\lim_{t\to\infty}P(t)=\frac{E(Z_n)}{E[Z_n]+E[Y_n]}$$

证明 每当系统打开时就说发生一次更新。对 t 之前(或 t 时)最后一次更新发生的时刻取条件得

$$\begin{aligned}P(t)&=P\{\text{在时刻}\,t\,\text{开着}\mid S_{N(t)}=0\}P\{S_{N(t)}=0\}\\&\quad+\int_0^{+\infty}P\{\text{在时刻}\,t\,\text{开着}\mid S_{N(t)}=y\}dF_{S_{N(t)}}(y)\end{aligned}$$

今有

$$\begin{aligned}P\{\text{在时刻}\,t\,\text{开着}\mid S_{N(t)}=0\}&=P\{Z_1>t\mid Z_1+Y_1>t\}\\&=\overline{H}(t)/\overline{F}(t)\end{aligned}$$

且对 $y>t$

$$\begin{aligned}P\{\text{在时刻}\,t\,\text{开着}\mid S_{N(t)}=y\}&=P\{Z>t-y\mid Z+Y>t-y\}\\&=\overline{H}(t-y)/\overline{F}(t-y)\end{aligned}$$

因此，利用引理 3.4.3 得

$$P(t)=\overline{H}(t)+\int_0^t\overline{H}(t-y)dm(y)$$

其中 $m(y)=\sum_{n=1}^{\infty}F_n(y)$。$\overline{H}(t)$ 显然是非负不增的，且 $\int_0^{+\infty}\overline{H}(t)dt=E[Z]<\infty$。因为后一式也蕴含了 $t\to\infty$ 时，$\overline{H}(t)\to 0$，所以只要利用关键更新定理我们有

$$P(t) \to \frac{\int_0^{+\infty} \overline{H}(t)dt}{\mu_F} = \frac{E[Z_n]}{E[Z_n] + E[Y_n]}$$

若令 $Q(t) = P\{$在时刻 t 时关着$\} = 1 - P(t)$，则

$$Q(t) \to \frac{E[Y]}{E[Z] + E[Y]}$$

我们注意到，系统最初是开着的这一事实在极限中不起什么作用。

定理 3.4.4 是十分重要的，因为许多系统能以交错过程为模型。例如，考虑一个更新过程，以 $Y(t)$ 记以 t 直到下一次更新的时间，而以 $A(t)$ 记$[0,t]$内最后一次更新之后的时间，即

$$Y(t) = S_{N(t)+1} - t$$
$$A(t) = t - S_{N(t)}$$

称 $Y(t)$ 是时刻 t 的过剩或剩余寿命，而称 $A(t)$ 是时刻 t 的年龄。若假设更新过程是将一个部件投入使用而一旦失效即更换所产生的，则 $A(t)$ 表示在时刻 t 所使用的部件的年龄而 $Y(t)$ 表示它的剩余寿命。

假设我们想推导出 $P\{A(t) \leqslant x\}$。为此将一个开—关的循环对应于一个更新区间，且若在 t 时的年龄小于或等于 x，就说系统在时刻 t "开着"。换言之，在一个更新区间的前 x 时间内系统"开着"而其余时间"关着"。那么，若更新分布不是格点的，则由定理 3.4.4 得

$$\begin{aligned}\lim_{t\to\infty} P\{A(t) \leqslant x\} &= E[\min(X,x)]/E[X] \\ &= \int_0^{+\infty} P\{\min(X,x) > y\}dy/E[X] \\ &= \int_0^x \overline{F}(y)dy/\mu\end{aligned}$$

类似地为了得到 $P\{Y(t) \leqslant x\}$ 的极限值，我们说在一个更新循环的最后 x 时间内系统"关着"而其余时间"开着"。于是在一次循环中"关着"的时间是 $\min(x, X)$，从而

$$\begin{aligned}\lim_{t\to\infty} P\{Y(t) \leqslant x\} &= \lim_{t\to\infty} P\{\text{在时刻 } t \text{ 关着}\} \\ &= E[\min(x,X)]/E[X]\end{aligned}$$

$$= \int_0^x \overline{F}(y)dy/\mu$$

于是合起来我们证明了下列

命题 3.4.5 若来到间隔的分布不是格点的，且 $\mu<\infty$，则

$$\lim_{t\to\infty} P\{Y(t)\leqslant x\} = \lim_{t\to\infty} P\{A(t)\leqslant x\} = \int_0^x \overline{F}(y)dy/\mu$$

注记。为了理解剩余寿命与年龄的极限分布为什么是相同的，考虑一个已运行很长时间的过程；例如，假设它起始于 $t=-\infty$。如果我们把时间倒过去往回观察，那么相继事件之间的间隔仍然独立且具有分布 F。因此，往回观察看到的是一个分布相同的更新过程。但往回观察时，在时刻 t 的剩余寿命正是原过程在时刻 t 的年龄。我们将发现把时间倒过去往回看的技巧在第四章与第五章中研究马尔可夫链时是十分有价值的。(使剩余寿命与年龄的分布发生关系的另一方法参见习题 11。)

另一有趣的随机变量是 $X_{N(t)+1}=S_{N(t)+1}-S_{N(t)}$，或等价地

$$X_{N(t)+1} = A(t) + Y(t)$$

于是 $X_{N(t)+1}$ 表示包含点 t 的更新区间的长度。在习题 3 中我们证明

$$P\{X_{N(t)+1} > x\} \geqslant \overline{F}(x)$$

即是对任何 x，含点 t 的更新区间之长大于 x 较之于一个普通的更新区间之长大于 x 更加可能发生。这个结论，初看起来似乎令人吃惊，它以检查悖论著称。

现在我们用交错更新理论来得到 $X_{N(t)+1}$ 的极限分布。

仍设一个开一关循环对应于一个更新区间，且如果整个循环时间大于 x 就说整个循环时间都是开着的，否则开着的时间就是零。即此系统在一循环期中或全开着(当更新区间长度大于 x 时)或全关着。现在

$$P\{X_{N(t)+1} > x\} = P\{\text{包含 } t \text{ 的更新区间长度} > x\}$$
$$= P\{\text{在时刻 } t \text{ 开着}\}$$

于是由定理 3.4.4，若 F 不是格点的，得

$$\lim_{t\to\infty}P\{X_{N(t)+1}>x\}=\frac{E[\text{循环中的开时}]}{\mu}$$

$$=E[X\mid X>x]\overline{F}(x)/\mu$$

$$=\int_x^{+\infty}ydF(y)/\mu$$

或等价地有

(3.4.2) $$\lim_{t\to\infty}P\{X_{N(t)+1}\leqslant x\}=\int_0^x ydF(y)/\mu$$

注记　为了更好地理解检查悖论，作如下推理：因为直线被更新区间所复盖，较长的区间(相对于较短的区间而言)复盖点 t 的可能性不是更大吗？事实上，在极限中(当 $t\to\infty$)确实是这样，长为 y 的区间与长为 1 的区间相比，复盖 t 的可能性，前者 y 倍于后者。如果真是这样，那么包含点 t 的区间的密度，记作 $g(y)$，就会是 $g(y)dy=ydF(y)/\mu$(因为 $dF(y)$ 是任意一个区间长为 y 的概率，而 y/μ 是此区间含点 t 的条件概率)。但由(3.4.2)看到这实际上正是极限密度。

考虑下列，它又一次说明了交错更新过程的多种用法。

例 3.4(a)　存储论一例。假设顾客按一更新过程来到一家出售单一品种商品的商店，来到间隔分布 F 是非格点的。顾客的需求量假定是独立的，具有共同的分布 G。商店使用如下的 (s,S) 定货策略：若在为一顾客服务之后存货量低于 s，则立即定货使之达到 S，否则不定货。于是若为一顾客服务之后存货量为 x，则定货量是

$$\begin{array}{ll} S-x, & \text{若 } x<s \\ 0, & \text{若 } x\geqslant s \end{array}$$

这里假定定货瞬时间即被补足。

以 $X(t)$ 记时刻 t 的存货量，我们想求 $\lim_{t\to\infty}P\{X(t)\geqslant x\}$。设 $X(0)=S$，如果在货存量至少为 x 时我们说系统是“开”的，否则是“关”的，那么这正是一个交错更新过程。因此，从定理 3.4.4 得

$$\lim_{t\to\infty}P\{X(t)\geqslant x\}=\frac{E[\text{在一个循环中存货量}\geqslant x\text{ 的时间}]}{E[\text{一个循环的时间}]}$$

现在如果我们以 Y_1，Y_2，…记相继来到的顾客的需求量且设

$$N_x = \min\{n : Y_1 + \cdots + Y_n > S - x\} \tag{3.4.3}$$

那么正是这一个循环中的第 N_x 个顾客使得存货量降到 x 之下，而正是第 N_s 个顾客结束这一循环。因此，以 X_i（$i \geqslant 1$）记顾客的来到间隔，则

$$\text{在一个循环中“开”的时间} = \sum_{i=1}^{N_x} X_i$$

$$\text{一个循环的时间} = \sum_{i=1}^{N_s} X_i$$

假定来到间隔与相继的需求量独立，于是取期望得

$$\lim_{t\to\infty} P\{X(t) \geqslant x\} = \frac{E\left[\sum_{i=1}^{N_x} X_i\right]}{E\left[\sum_{i=1}^{N_s} X_i\right]} = \frac{E[N_x]}{E[N_s]} \tag{3.4.4}$$

然而，因为 Y_i，$i \geqslant 1$，独立同分布，由(3.4.3)我们可把 $N_x - 1$ 解释为一个来到间隔为 Y_i，$i \geqslant 1$，的更新过程到时刻 $S-x$ 为止更新的次数。因此，

$$E[N_x] = m_G(S-x) + 1$$

$$E[N_s] = m_G(S-s) + 1$$

其中 G 是顾客需求量的分布且

$$m_G(t) = \sum_{n=1}^{\infty} G_n(t)$$

因此，从(3.4.4)得到

$$\lim_{t\to\infty} P\{X(t) \geqslant x\} = \frac{1 + m_G(S-x)}{1 + m_G(S-s)}, \quad x \leqslant S$$

3.4.2 极限平均剩余寿命与 $m(t)$ 的展开式

让我们从计算非格点的更新过程的平均剩余寿命开始，关于 $S_{N(t)}$ 取条件得（据引理 3.4.3）

$$E[Y(t)] = E[Y(t) \mid S_{N(t)} = 0]\overline{F}(t) + \int_0^t E[Y(t) \mid S_{N(t)} = y]\overline{F}(t-y)\,dm(y)$$

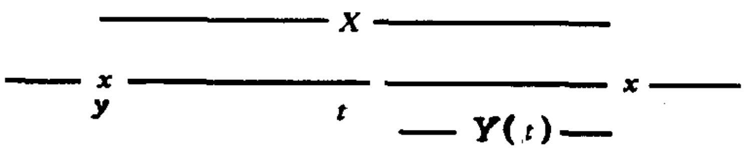


图 3.4.1 $S_{N(t)} = y$；$x =$ 更新

现在，

$$E[Y(t)|S_{N(t)}=0]=E[X-t|X>t]$$

$$E[Y(t)|S_{N(t)}=y]=E[X-(t-y)|X>t-y]$$

这里上式成立是因为$S_{N(t)}=y$意味着在y有一更新且下一个来到的时间(记为X)要大于$t-y$(见图3.4.1)。因此，

$$E[Y(t)]=E[X\text{-}t|X>t]\overline{F}(t)+\int_0^t E[X-(t-y)|X>t-y]\overline{F}(t-y)dm(y)$$

现在可以证明，倘若$E[X^2]<\infty$，函数$h(t)=E[X-t|X>t]\overline{F}(t)$直接黎曼可积，从而由关键更新定理得

$$\begin{aligned}E[Y(t)]&\to\int_0^\infty E[X-t|X>t]\overline{F}(t)dt/\mu\\&=\int_0^\infty\int_t^\infty(x-t)dF(x)dt/\mu\\&=\int_0^\infty\int_0^x(x-t)dtdF(x)/\mu\quad\text{（交换积分次序）}\\&=\int_0^\infty x^2dF(x)/2\mu\\&=E[X^2]/2\mu\end{aligned}$$

于是证得下列

命题 3.4.6

如果来到间隔分布不是格点的，且$E[X^2]<\infty$，则

$$\lim_{t\to\infty}E[Y(t)]=E[X^2]/2\mu$$

现在$S_{N(t)+1}$，即t后的第一个更新时刻能表示为

$$S_{N(t)+1}=t+Y(t)$$

取期望且利用系3.3.3，我们有

$$\mu(m(t)+1)=t+E[Y(t)]$$

或

$$m(t)-\frac{t}{\mu}=\frac{E[Y(t)]}{\mu}-1$$

因此，从命题3.4.6得下列

系 3.4.7

若 $E[X^2]<\infty$ 且 F 为非格点的，则 $t\to\infty$ 时，

$$m(t)-\frac{t}{\mu}\to\frac{E[X^2]}{2\mu^2}-1$$

3.4.3 年龄相依的分支过程

假设一生物在它的生命结束时按概率分布$\{P_j,\ j=0,\ 1,\ 2,\ \cdots\}$产生出随机个后代。进一步假定所有的后代彼此独立地活动且依同一概率分布$\{P_j\}$产生各自的后代。最后，让我们假定这些生物的寿命是独立的随机变量，具有同样的分布 F。

以 $X(t)$ 记在时刻 t 活着的生物个数。此随机过程$\{X(t),t\geqslant 0\}$称为年龄相依的分支过程。我们关心的是确定 $m=\sum_{j=0}^{\infty}jP_j>1$时 $M(t)=E[X(t)]$ 的渐近式。

定理 3.4.8　若 $X_0=1$，$m>1$，且 F 不是格点的，则 $t\to\infty$时，

$$e^{-\alpha t}M(t)\to\frac{m-1}{m^2\alpha\int_0^{+\infty}xe^{-\alpha x}dF(x)}$$

其中 α 是满足下式的唯一正数

$$\int_0^{+\infty}e^{-\alpha x}dF(x)=\frac{1}{m}$$

证明　关于 T_1（即最初始生物的寿命）取条件，我们得

$$M(t)=\int_0^{+\infty}E[X(t)|T_1=s]dF(s)$$

然而，

$$E[X(t)|T_1=s]=\begin{cases}1, & 若\ s>t\\ m\cdot M(t-s), & 若\ s\leqslant t\end{cases}\tag{3.4.5}$$

为了理解(3.4.5)为什么成立，假设 $T_1=s$，$s\leqslant t$，且进一步假设此生物有 j 个后代，则在时刻 t 活着的生物个数可写成 $Y_1+\cdots+Y_j$，其中 Y_i 是在时刻 t 活着的第 i 个后代的子孙(包括它自己)个数。显然，Y_1，$\cdots Y_j$ 独立，具有与 $X(t-s)$ 同样的分布。故$E(Y_1+\cdots+Y_j)=jM(t-s)$；取 $jM(t-s)$的期望(对 j)得到(3.4.5)。

于是从上面我们得到

(3.4.6) $$M(t) = \overline{F}(t) + m\int_0^t M(s-t)dF(s)$$

现在以 α 记满足下式的唯一正数：

$$\int_0^{+\infty} e^{-\alpha y}dF(y) = \frac{1}{m}$$

且定义分布 G 为

$$G(s) = m\int_0^s e^{-\alpha y}dF(y), \qquad 0 \leqslant s < \infty$$

在(3.4.6)式两边乘以 $e^{-\alpha t}$ 且利用 $dG(s)=me^{-\alpha s}dF(s)$，我们得到

(3.4.7) $$e^{-\alpha t}M(t) = e^{-\alpha t}\overline{F}(t) + \int_0^t e^{-\alpha(t-s)}M(t-s)dG(s)$$

令 $f(t)=e^{-\alpha t}M(t)$ 及 $h(t)=e^{-\alpha t}\overline{F}(t)$，运用卷积符号从(3.4.7)我们有

$$\begin{aligned}
f &= h+f*G \\
&= h+G*f \\
&= h+G*(h+G*f) \\
&= h+G*h+G_2*f \\
&= h+G*h+G_2*(h+G*f) \\
&= h+G*h+G_2*h+G_3*f \\
&\vdots \\
&= h+G*h+G_2*h+\cdots+G_n*h+G_{n+1}*f
\end{aligned}$$

由于 G 是某个非负随机变量的分布，故得 $n\to\infty$ 时，$G_n(t)\to 0$（为什么?）。从而在上式中令 $n\to\infty$ 得

$$\begin{aligned}
f &= h + h*\sum_{i=1}^{\infty}G_i \\
&= h + h*m_G
\end{aligned}$$

或

$$f(t) = h(t) + \int_0^t h(t-s)dm_G(s)$$

可证明 $h(t)$ 是直接黎曼可积的，故由关键更新定理

(3.4.8) $$f(t) \to \frac{\int_0^{+\infty} h(t)dt}{\mu_G} = \frac{\int_0^{+\infty} e^{-\alpha t}\overline{F}(t)dt}{\int_0^{+\infty} xdG(x)}$$

而

(3.4.9) $$\int_0^{+\infty} e^{-\alpha t}\overline{F}(t)dt = \int_0^{+\infty} e^{-\alpha t}\int_t^{+\infty} dF(x)dt$$

$$
\begin{aligned}
&= \int_0^{+\infty}\int_0^x e^{-\alpha t}dtdF(x) \\
&= \frac{1}{\alpha}\int_0^{+\infty}(1 - e^{-\alpha x})dF(x) \\
&= \frac{1}{\alpha}(1 - \frac{1}{m}) \quad （由 \alpha 的定义）
\end{aligned}
$$

又

(3.4.10)
$$\int_0^{\infty} xdG(x) = m\int_0^{+\infty} xe^{-\alpha x}dF(x)$$

于是从（3.4.8），（3.4.9）及（3.4.10）得到 $t\to\infty$ 时

$$e^{-\alpha t}M(t) \to \frac{m-1}{m^2\alpha\int_0^{+\infty} xe^{-\alpha x}dF(x)}$$

3.5 延迟更新过程

我们常常考虑这样一种计数过程，它的第一个来到间隔与其余的有不同的分布。例如，我们可能以在某时刻 $t>0$ 开始观察一个更新过程。若在时刻 t 未发生更新，则我们首次观察到更新所必须等待的时间的分布不同于其余来到间隔的分布。

正式地，设 $\{X_n, n=1,2,\cdots\}$ 为一列独立非负随机变量，X_1 具有分布 G，而 X_n 具有分布 F，$n>1$。令 $S_0=0$，$S_n=\sum_{i=1}^{n} X_i$，$n\geqslant 1$，且定义

$$N_D(t)=\sup\{n: S_n\leqslant t\}$$

定义 随机过程 $\{N_D(t), t\geqslant 0\}$ 称为一般或延迟更新过程。

当 $G=F$ 时，自然这就是通常的更新过程。如同通常的情形，我们有

$$
\begin{aligned}
P\{N_D(t)=n\} &= P\{S_n\leqslant t\} - P\{S_{n+1}\leqslant t\} \\
&= G * F_{n-1}(t) - G * F_n(t)
\end{aligned}
$$

令

$$m_D(t)=E[N_D(t)]$$

则容易证明

(3.5.1) $$m_D(t)=\sum_{n=1}^{\infty}G*F_{n-1}(t)$$

且对(3.5.1)取变换得

(3.5.2) $$\widetilde{m}_D(s)=\frac{\widetilde{G}(s)}{1-\widetilde{F}(s)}$$

利用通常更新过程的相应结果，容易证得延迟过程的类似的极限定理。我们将下列命题的证明留给读者。

令 $\mu=\int_0^{+\infty}xdF(x)$ 。

命题 3.5.1

(i) 以概率 1，$t\to\infty$时

$$\frac{N_D(t)}{t}\to\frac{1}{\mu}$$

(ii) $t\to\infty$时，$\frac{m_D(t)}{t}\to\frac{1}{\mu}$。

(iii) 若 F 不是格点的，则 $t\to\infty$时

$$m_D(t+a)-m_D(t)\to\frac{a}{\mu}$$

(iv) 若 F 与 G 是格点的，周期为 d，则 $n\to\infty$时

E[在时刻 nd 的更新次数] $\to\frac{d}{\mu}$

(v) 若 F 不是格点的，$\mu<\infty$，及 h 直接黎曼可积，则

$$\int_0^{+\infty}h(t-x)dm_D(x)\to\int_0^{+\infty}h(t)dt/\mu$$

例 3.5(a) 假设观察了一列独立同分布的离散随机变量 $X_1,X_2,\cdots$，且假设我们跟踪一个花样（即给定的一列试验结果）出现的次数。如设花样是 $x_1, x_2, \cdots, x_k$，我们说在时刻 n 这花样发生，如果 $X_n=x_k$, $X_{n-1}=x_{k-1}$, $\cdots$, $X_{n-k+1}=x_1$。例如，若花样是 0，1，0，1，而序列是$(X_1,X_2,\cdots)=$ (1，0，1，0，1，0，1，1，1，0，1，0，1，…)，则在时刻 5，7，13 这花样发生。若我们以 $N(n)$记到时刻 n 发生这花样的次数，则$\{N(n),n\geqslant 1\}$是一个延迟更新过程。首次更新时间的分布是这花样首次发生的时间的分布；而后面的来到间隔分布是这花样的各次复制之间的时间分布。

假定我们想确定此花样发生的速率。由延迟更新过程的强大数定律（定

理 3.5.1 的(i))，此速率将等于花样之间的平均时间的倒数。但由布莱克威尔定理（定理 3.5.1 的 (iv))，这正是在时刻 n 有一个更新发生的极限概率，即

$$(E[\text{花样之间的时间}])^{-1}=\lim_{n\to\infty}P\{\text{在时刻 } n \text{ 花样发生}\}$$
$$=\prod_{i=1}^{k}P\{X=x_i\}$$

因此花样发生的速率是 $\prod_{i=1}^{k}P\{X=x_i\}$，而花样之间的平均时间是$\left(\prod_{i=1}^{k}P\{X=x_i\}\right)^{-1}$。

例 3.5(b) 一个系统由 n 个独立部件组成，各部件的活动是一个指数交替更新过程。更具体地，部件 i，$i=1,2,\cdots n$，处于正常状态的时间是均值为 λ_i 的指数变量，而后失效，在恢复正常重新开始之前滞留于失效状态的时间是均值为 μ_i 的指数变量。

假设在任何时刻至少有一部件处于正常状态便称此系统这时在工作(这样的系统称为并联系统)。如果我们以 $N(t)$ 记此系统在$[0,t]$内变成不工作（即发生故障）的次数，则$\{N(t),t\geqslant 0\}$是一个延迟更新过程。

我们要计算系统发生故障之间的平均时间。为此我们先看看当 t 很大而 h 很小时在 $(t,t+h)$ 中发生一次故障的概率，在 $(t,t+h)$ 中发生一次故障的一种情形是在时刻 t 恰有一个部件正常而其它的全部失效，而后此部件损坏。因为一切其它可能的情形合在一起的概率显然是 $o(h)$，可见

$$\lim_{t\to\infty}P\{\text{在}(t,t+h)\text{中出故障}\}=\sum_{i=1}^{n}\left\{\frac{\lambda_i}{\lambda_i+\mu_i}\prod_{j\neq i}\frac{\mu_j}{\lambda_j+\mu_j}\right\}\frac{1}{\lambda_i}h+o(h)$$

但由布莱克威尔定理上式正是故障之间的平均时间的倒数的 h 倍，所以令 $h\to 0$ 即得

$$E[\text{故障之间的时间}]=\left(\prod_{j=1}^{n}\frac{\mu_j}{\lambda_j+\mu_j}\sum_{i=1}^{n}\frac{1}{\mu_i}\right)^{-1}$$

因为一个故障期的平均长度是$\left(\sum_{j=1}^{n}1/\mu_i\right)^{-1}$，所以我们能算得一个工作期的平均长度为

$$E[\text{工作期的长度}]$$
$$=E[\text{两次故障之间的时间}]-\left(\sum_{i=1}^{n}\frac{1}{\mu_i}\right)^{-1}$$

$$=\frac{1-\prod_{j=1}^{n}\frac{\mu_j}{\lambda_j+\mu_j}}{\prod_{j=1}^{n}\frac{\mu_j}{\lambda_j+\mu_j}\sum_{i=1}^{n}\frac{1}{\mu_i}}$$

为了检验上式，注意到这系统可以被看作一个延迟的交错更新过程，其处于不工作状态的极限概率是

$$\lim_{t\to\infty}P\{\text{在 } t \text{ 时系统故障}\}=\prod_{j=1}^{n}\frac{\mu_j}{\lambda_j+\mu_j}$$

现在我们可以验证上式实际上等于一个故障期的平均长度除以一个循环(或故障之间的时间)的平均长度。

按照通常更新过程中采用的同样的证明方法可得，在 t 之前(或在时刻 t)的最后一次更新时刻的分布为

(3.5.3) $$P\{S_{N(t)}\leqslant s\}=\overline{G}(t)+\int_0^s\overline{F}(t-y)dm_D(y)$$

当 $\mu<\infty$时，分布函数

$$F_e(x)=\int_0^x\overline{F}(y)dy/\mu,\quad x\geqslant 0$$

称为 F 的平衡分布，其拉普拉斯变换为

(3.5.4) $$\begin{aligned}\widetilde{F}_e(s)&=\int_0^{+\infty}e^{-sx}dF_e(x)\\&=\int_0^{+\infty}e^{-sx}\int_x^{\infty}dF(y)dx/\mu\\&=\int_0^{\infty}\int_0^{y}e^{-sx}dxdF(y)/\mu\\&=\frac{1}{s\mu}\int_0^{\infty}(1-e^{-sy})dF(y)\\&=\frac{1-\widetilde{F}(s)}{\mu s}\end{aligned}$$

$G=F_e$ 的延迟更新过程称为平衡更新过程，它是极其重要的，因为假设在时刻 t 我们开始观察一个更新过程。那么我们所观察的过程是一个延迟更新过程，其初始分布是 $Y(t)$的分布。于是当 t 很大时，从命题 3.4.5 得所观察的过程是平衡更新过程。此过程中的平衡性在下一定理中证明。

以 $Y_D(t)$记一个延迟更新过程在时刻 t 的剩余寿命。

定理 3.5.2 对于平衡更新过程有：

(i) $m_D(t)=t/\mu$；

(ii) $P\{Y_D(t)\leqslant x\}=F_e(x)$，对一切 $t\geqslant 0$ 成立；

(iii) $\{N_D(t),t\geqslant 0\}$具有平稳增量。

证明 (i) 从(3.5.2)与(3.5.4)我们有

$$\widetilde{m}_D(s)=\frac{1}{\mu s}$$

然而，简单的计算表明 $1/\mu s$ 是函数 $h(t)=t/\mu$ 的拉普拉斯变换，于是由变换的唯一性可得

$$m_D(t)=t/\mu$$

(ii) 对于一个延迟更新过程，关于 $S_{N}(t)$取条件，利用(3.5.3)可得

$$P\{Y_D(t)>x\}=P\{Y_D(t)>x\,|\,S_{N(t)}=0\}\overline{G}(t)$$
$$+\int_0^t P\{Y_D(t)>x\,|\,S_{N(t)}=s\}\overline{F}(t-s)dm_D(s)$$

今有

$$P\{Y_D(t)>x\,|\,S_{N(t)}=0\}=P\{X_1>t+x\,|\,X_1>t\}$$
$$=\frac{\overline{G}(t+x)}{\overline{G}(t)}$$
$$P\{Y_D(t)>x\,|\,S_{N(t)}=s\}=P\{X>t+x-s\,|\,X>t-s\}$$
$$=\frac{\overline{F}(t+x-s)}{\overline{F}(t-s)}$$

因此，

$$P\{Y_D(t)>x\}=\overline{G}(t+x)+\int_0^t\overline{F}(t+x-s)dm_D(s)$$

令 $G=F_e$ 且利用(i)的结论得

$$\begin{aligned}P\{Y_D(t)>x\}&=\overline{F}_e(t+x)+\int_0^t\overline{F}(t+x-s)ds/\mu\\&=\overline{F}_e(t+x)+\int_x^{x+t}\overline{F}(y)dy/\mu\\&=\overline{F}_e(x)\end{aligned}$$

(iii) 为证(iii)，注意到 $N_D(t+s)-N_D(s)$可解释为一个延迟更新过程在 t 时间内更新的次数，其初始分布是 $Y_D(s)$的分布。这样就从(ii)得(iii)的结论。

3.6 更新酬劳过程

大量的概率模型是以下模型的特殊情形。考虑一个更新过程$\{N(t),t\geqslant 0\}$,其来到间隔 X_n, $n\geqslant 1$,具有分布 F。且假设每次一个更新发生,我们收到一份酬劳。以 R_n 记在第 n 次更新时刻所获得的酬劳。我们将假定 R_n, $n\geqslant 1$,独立同分布。但我们允许 R_n 可以(通常也将)依赖于 X_n,第 n 个更新区间的长度。所以我们假定(X_n,R_n),$n\geqslant 1$,是独立同分布的。令

$$R(t)=\sum_{n=1}^{N(t)} R_n$$

则 $R(t)$ 表示到时间 t 所得的全部酬劳。设

$$E[R]=E[R_n],\quad E[X]=E[X_n]$$

定理 3.6.1 若 $E[R]<\infty$ 及 $E[X]<\infty$,则

(i) 以概率 1,$t\to\infty$ 时

$$\frac{R(t)}{t}\to\frac{E[R]}{E[X]}$$

(ii) $t\to\infty$, $\dfrac{E[R(t)]}{t}\to\dfrac{E[R]}{E[X]}$

证明 为证(i)写

$$\frac{R(t)}{t}=\frac{\sum_{n=1}^{N(t)} R_n}{t}$$

$$=\left(\frac{\sum_{n=1}^{N(t)} R_n}{N(t)}\right)\left(\frac{N(t)}{t}\right)$$

由强大数定律可得,$t\to\infty$ 时

$$\frac{\sum_{n=1}^{N(t)} R_n}{N(t)}\to E[R]$$

而由更新过程的强极限律,$t\to\infty$ 时

$$\frac{N(t)}{t}\to\frac{1}{E[X]}$$

于是(i)得证。

为证明(ii)我们首先注意到 $N(t)+1$ 是序列 $X_1,X_2,\cdots$ 的一个停时，也是 $R_1,R_2,\cdots$ 的一个停时(为什么?)。于是由瓦尔德等式，

$$E\left[\sum_{i=1}^{N(t)}R_i\right]=E\left[\sum_{i=1}^{N(t)+1}R_i\right]-E[R_{N(t)+1}]$$

$$=(m(t)+1)E[R]-E[R_{N(t)+1}]$$

从而

$$\frac{E[R(t)]}{t}=\frac{m(t)+1}{t}E[R]-\frac{E[R_{N(t)+1}]}{t}$$

若能证明 $t\to\infty$ 时 $E[R_{N(t)+1}]/t\to 0$，则由基本更新定理可得所求证的结论。为此，令 $g(t)=E[R_{N(t)+1}]$。关于 $S_{N(t)}$ 取条件得

$$g(t)=E[R_{N(t)+1}|S_{N(t)}=0]\overline{F}(t)$$

$$+\int_0^t E[R_{N(t)+1}|S_{N(t)}=s]\overline{F}(t-s)dm(s)$$

然而，

$$E[R_{N(t)+1}|S_{N(t)}=0]=E[R_1|X_1>t]$$

$$E[R_{N(t)+1}|S_{N(t)}=s]=E[R_n|X_n>t-s]$$

所以

$$g(t)=E[R_1|X_1>t]\overline{F}(t)+\int_0^t E[R_n|X_n>t-s]\overline{F}(t-s)dm(s) \tag{3.6.1}$$

现在令

$$h(t)=E[R_1|X_1>t]\overline{F}(t)=\int_t^{+\infty}E[R_1|X_1=x]dF(x)$$

且注意到因为

$$E|R_1|=\int_0^{+\infty}E[|R_1||X_1=x]dF(x)<\infty$$

得

$$t\to\infty\text{时}，h(t)\to 0 \quad \text{且} \quad h(t)\leqslant E|R_1|，\quad \text{对一切}\ t\ \text{成立}，$$

于是可选取 T 使 $|h(t)|<\varepsilon$ 对任何 $t\geqslant T$ 成立。因此从(3.6.1)，由基本更新定理得

$$\frac{|g(t)|}{t}\leqslant\frac{|h(t)|}{t}+\int_0^{t-T}\frac{|h(t-x)|}{t}dm(x)+\int_{t-T}^{t}\frac{|h(t-x)|dm(x)}{t}$$

$$\leqslant\frac{\varepsilon}{t}+\frac{\varepsilon m(t-T)}{t}+E|R_1|\frac{m(t)-m(t-T)}{t}$$

$$\rightarrow \frac{\varepsilon}{E[X]}, \quad 当 t\rightarrow\infty$$

因 ε 可任意小而得 $g(t)/t\rightarrow 0$，所要的结论得证。

注记。若每当一更新发生我们就说完成了一个循环，那么此定理说长时间之后的平均酬劳（或酬劳的期望）恰是一个循环中所得的酬劳的期望除以一个循环的时间的期望。

证明此定理时会受到诱惑去断言 $E[R_{N(t)+1}]=E[R]$，从而显然地得到 $E[R_{N(t)+1}]/t$ 收敛于零。然而 $R_{N(t)+1}$ 与 $X_{N(t)+1}$ 有关，且 $X_{N(t)+1}$ 是包含点 t 的更新区间的长度，因为较长的更新区间有更多的机会包含点 t，（直观上）$X_{N(t)+1}$ 趋向于比一个通常的更新区间长度更大的值（见习题 3），所以 $R_{N(t)+1}$ 的分布不同于 R_1 的分布。

到目前为止我们都假定在更新循环一结束时就立刻得到酬劳。但这并不是本质的，若酬劳是在更新循环期间逐步地获得的，定理 3.6.1 仍然正确。为了明瞭这点，以 $R(t)$ 记到 t 时所得的酬劳，且先假设一切报酬是非负的。于是

$$\frac{\sum_{n=1}^{N(t)} R_n}{t} \leqslant \frac{R(t)}{t} \leqslant \frac{\sum_{n=1}^{N(t)} R_n}{t} + \frac{R_{N(t)+1}}{t}$$

由于

$$\frac{E[R_{N(t)+1}]}{t} \rightarrow 0$$

得到定理 3.6.1 的（ii）。注意到，由证明中的论证 $\sum_{n=1}^{N(t)} R_n/t$ 与 $\sum_{n=1}^{N(t)+1} R_n/t$ 都收敛于 $E[R]/E[X]$，从而得到定理 3.6.1 的(i)。当酬劳非正时类似的论证也成立，将酬劳分为正部与负部，并对每一部分各自运用以上论证即得一般情形的结论。

例 3.6(a)　交错更新过程。对于一个交错更新过程（见 3.4.1 节）。假设系统开时以每单位时间为 1 的比率获得报酬（因此一次循环所得酬劳等于此循环中开的时间）。到 t 时所得的总的酬劳正是 $[0, t]$ 中开着的总时间，于是由定理 3.6.1 得，以概率 1

$$\frac{\text{在}[0,t]\text{中开着的时间}}{t} \to \frac{E[X]}{E[X]+E[Y]}$$

其中 X 是一个循环中开着的时间，而 Y 是关着的时间，再由定理 3.4.4 知，当一个循环的分布为非格点时此系统开着的极限概率等于长时间后开着的时间所占的比率。

例 3.6(b)　平均年龄与剩余寿命。 此 $A(t)$ 记一个更新过程在时刻 t 的年龄，假设我们感兴趣的是要计算

$$\lim_{t\to\infty}\int_0^t A(s)ds/t$$

为此假定在任何时刻我们以与该时刻更新过程的年龄相等的比率收钱，即在时刻 s 以比率 $A(s)$ 收到酬金，所以 $\int_0^t A(s)ds$ 表示到时刻 t 我们的总收益。因为每当一次更新发生，一切又从头开始，所以以概率 1

$$\frac{\int_0^t A(s)ds}{t} \to \frac{E[\text{一个更新循环中的酬劳}]}{E[\text{一个更新循环的时间}]}$$

现在由于进入一个更新循环时间 s 之后，更新过程的年龄正是 s，所以我们有

$$\text{一个更新循环中的酬劳} = \int_0^X sds = \frac{X^2}{2}$$

其中 X 是更新循环的时间。因此，以概率 1 有

$$\lim_{t\to\infty}\frac{\int_0^t A(s)ds}{t} = \frac{E[X^2]}{2E[X]}$$

类似地若以 $Y(t)$ 记时刻 t 的剩余寿命，那么假设我们以与该时刻的剩余寿命相等的比率获得酬劳便可计算平均剩余寿命。由定理 3.6.1 剩余寿命的平均值为

$$\lim_{t\to\infty}\int_0^t Y(s)ds/t = \frac{E[\text{一个更新循环中的酬劳}]}{E[X]}$$
$$= \frac{E[X^2]}{2E[X]}$$

于是年龄与剩余寿命的均值相等。(为什么这是所期望的?)

量 $X_{N(t)+1}=S_{N(t)+1}-S_{N(t)}$ 表示包含点 t 的更新区间的长度。因为它可以表示为

$$X_{N(t)+1}=A(t)+Y(t)$$

可见它的均值为

$$\lim_{t\to\infty}\int_0^t X_{N(s)+1}ds/t = \frac{E[X^2]}{E[X]}$$

因为

$$\frac{E[X^2]}{E[X]} \geqslant E[X]$$

(等式仅当 $\mathrm{Var}(X)=0$ 时成立)，可见 $X_{N(t)+1}$的均值大于 $E[X]$。(为什么这并不令人惊奇?)

例 3.6(c) 假设乘客按照一个更新过程来到一火车站，其平均来到间隔时间为 μ。每当有 N 个人在车站上等待时，就开出一辆火车。若每当有 n 个乘客等待时车站就以每单位时间 nc 元的比率开支费用，且每开出一辆火车要多开支 K 元，那么此车站每单位时间的平均费用是多少?

每当一辆火车开出我们就说完成了一次循环，那么上述过程是一更新酬劳过程。一次循环的平均长度是来到 N 个顾客所需的平均时间，而因为平均来到间隔时间是 μ，所以它等于

$$E[\text{循环的长度}] = N\mu$$

若以 X_n 记一次循环中第 n 个与第 $(n+1)$ 个来到之间的时间，则一个循环的平均费用可表示为

$$\begin{aligned} &E[\text{一次循环的费用}]\\ &= E[cX_1 + 2cX_2 + \cdots + (N-1)cX_{N-1}] + K\\ &= \frac{c\mu N(N-1)}{2} + K \end{aligned}$$

因此平均费用是

$$\frac{c(N-1)}{2} + \frac{K}{N\mu}$$

3.6.1 排队论应用一例

假设顾客遵循一个非格点的更新过程来到一个单个服务员的服务站。顾客一来到，若服务员空着则立即被服务，若服务员忙着则排队等候。假设顾客的服务时间独立同分布且与来到流独立。

以 X_1，X_2，…记顾客来到的间隔时间；以 Y_1，Y_2，…记顾客依次的服务时间。我们将假定

(3.6.1) $$E[Y_i] < E[X_i] < \infty$$

假设第一位顾客在时刻 0 来到，且以 $n(t)$ 记在时刻 t 系统中

顾客的人数。定义

$$L = \lim_{t\to\infty}\int_0^t n(s)ds/t$$

为了证明 L 以概率 1 存在且为常数，我们设想在时刻 s 所获酬劳率是 $n(s)$。若设一循环对应于一忙期的起点（即每当一个来客发现系统空着就开始一个新的循环），则容易看到每一次循环过程都重新开始。因 L 表示长时间后的平均酬劳，所以从定理 3.6.1 得

(3.6.2)
$$L = \frac{E[\text{一个循环中的酬劳}]}{E[\text{一个循环的时间}]} = \frac{E\left[\int_0^T n(s)ds\right]}{E[T]}$$

又以 W_i 记第 i 个顾客在此系统中花费的全部时间，且定义

$$W = \lim_{n\to\infty}\frac{W_1+\cdots+W_n}{n}$$

为证明 W 以概率 1 存在，设想我们在第 i 天收到酬劳 W_i。因为每次循环之后排队过程又重新开始，所以若以 N 记一次循环中受到服务的顾客数，那么 W 是一个更新过程的单位时间的平均酬劳，过程的循环时间是 N 而一个循环的酬劳是 $W_1+\cdots+W_N$，因而

(3.6.3)
$$W = \frac{E[\text{一个循环中的酬劳}]}{E[\text{一个循环的时间}]} = \frac{E\left[\sum_{i=1}^{N} W_1\right]}{E[N]}$$

我们应注意到，可以证明（见第七章命题 7.1.1）(3.6.1) 蕴含着 $E[N]<\infty$。

下述定理在排队论中是十分重要的。

定理 3.6.2 以 $\lambda = 1/E[X_i]$ 记来到率，则

$$L=\lambda W$$

证明 我们从一个循环的长度 T 与此循环中受到服务的顾客数 N 之间的关系着手。若在一个循环中 n 个顾客受到服务，则下一个循环在第 $(n+1)$ 个顾客来到时开始。因此，

$$T=\sum_{i=1}^{N} X_i$$

现在容易看到 N 是序列 X_1，X_2，…的一个停时，因为

$$N=n \Leftrightarrow X_1+X_2+\cdots+X_k<Y_1+\cdots+Y_k, \quad k=1, 2, \cdots, n-1$$

$$\text{且 } X_1+\cdots+X_n>Y_1+\cdots+Y_n$$

于是 $\{N=n\}$ 与 X_{n-1}，X_{n+2}，…独立，故由瓦尔德等式

$$E[T]=E[N]E[X]=E[N]/\lambda$$

所以由(3.6.2)与(3.6.3)得

$$L=\lambda W \frac{E\left[\int_0^T n(s)\ ds\right]}{E\left[\sum_{i=1}^{N} W_i\right]} \tag{3.6.4}$$

但设想各顾客在此系统中时以每单位时间 1 元的比率付款（所以第 i 个顾客所付总数正是 W_i），可见

$$\int_0^T n(s) ds=\sum_{i=1}^{N} W_i=\text{一个循环中所付总数}$$

因此从(3.6.4)证得结论。

注记

(1) 定理 3.6.2 的证明并不依赖于我们所假定的特殊排队模型。此证明可不作改变地适用于任何有如下特性的排队系统：该系统包含着在概率上过程重新开始的时刻，且这样的循环之间的平均是有限的。例如，若在我们的模型中假设有 k 个服务员，则可以证明平均循环时间有限的一个充分条件是

$$E[Y_i]<kE[X_i] \quad \text{及} \quad P\{Y_i<X_i\}>0$$

(2) 定理 3.6.2 说

在“系统”中(按时间的)平均人数

=λ·(一个顾客在“系统”中度过的平均时间)

以“队伍”替换“系统”，同样的证明得

在排队的平均人数

=λ·(一个顾客在排着的队伍中度过的平均时间)

或以“服务”替换“系统”我们就有

在一次服务中的平均人数 $=\lambda E[Y]$

3.7 再生过程

考虑状态空间为 $\{0,1,2,\cdots\}$ 的一个随机过程 $\{X(t),t\geqslant 0\}$，它具有这样的性质：存在着一些时刻，过程在这些时刻在概率上又重新开始，即假设以概率 1 存在一时刻 S_1 使得 S_1 之后过程的继续在概率上是从时刻 0 开始的全过程的复制。注意到这性质蕴含着还存在时刻 S_2，S_3，…它们与 S_1 有同样的性质。这样的随机过程称之为再生过程。

从上可得 $\{S_1,S_2,\cdots\}$ 构成一个更新过程的事件发生的时刻。每当发生一次更新，我们就说一个循环完成。以 $N(t)=\max\{n:S_n\leqslant t\}$ 记到时刻 t 循环的个数。

下列重要定理的证明进一步显示了关键更新定理的威力。

定理 3.7.1 若一个循环的分布 F 在某区间上具有密度，且若 $E[S_1]<\infty$，则

$$P_j=\lim_{t\to\infty}P\{X(t)=j\}=\frac{E[\text{一个循环中处于状态 } j \text{ 的时间}]}{E[\text{一个循环的时间}]}$$

证明 令 $P(t)=P\{X(t)=j\}$。对 t 之前最后一个循环的完成时刻取条件得

$$P(t)=P\{X(t)=j|S_{N(t)}=0\}\overline{F}(t)+\int_0^t P\{X(t)=j|S_{N(t)}=s\}\overline{F}(t-s)dm(s)$$

今有

$$P\{X(t)=j|S_{N(t)}=0\}=P\{X(t)=j|S_1>t\}$$
$$P\{X(t)=j|S_{N(t)}=s\}=P\{X(t-s)=j|S_1>t-s\}$$

于是

$$P(t)=P\{X(t)=j,S_1>t\}+\int_0^t P\{X(t-s)=j,S_1>t-s\}dm(s)$$

由于可以证明 $h(t)\equiv P\{X(t)=j,S_1>t\}$ 是直接黎曼可积的，所以由关键更新定理可得

$$P(t)\to\int_0^{+\infty}P\{X(t)=j,S_1>t\}dt/E[S_1]$$

今设

$$I(t)=\begin{cases}1, & 当 X(t)=j,S_1>t \\ 0, & 其它\end{cases}$$

则$\int_0^{+\infty} I(t)dt$代表着首次循环中$X(t)=j$的时间。因为

$$E\left[\int_0^{+\infty} I(t)dt\right]=\int_0^{+\infty} E[I(t)]dt$$

$$=\int_0^{\infty} P\{X(t)=j,S_1>t\}dt$$

故结论成立。

例 3.7(a)　顾客按更新过程来到的排队模型。顾客按更新过程来到的排队过程(如 3.6 节中的那样)多数是再生过程,每当一个来到的顾客发现系统是空着的,一个循环便开始。例如在顾客按更新过程来到的单服务员的排队系统中,时刻 t 系统中的人数 $X(t)$ 构成一个再生过程,倘若第一个顾客在 $t=0$ 时来到(否则它是一个延迟的再生过程,而定理 3.7.1 仍然有效)。

从更新酬劳过程的理论知 P_j 也等于长时间后 $X(t)=j$ 所占的时间的比率。事实上,我们有下列。

命题 3.7.2　对于 $E[S_1]<\infty$ 的再生过程,以概率 1

$$\lim_{t\to\infty}\frac{[(0,t)中处于 j 的时间]}{t}=\frac{E[一个循环中处于 j 的时间]}{E[一个循环的时间]}$$

证明　假设每当过程处于状态 j 时以比率 1 获得酬劳。这样生成一个更新酬劳过程,且从定理 3.6.1 直接得证本命题成立。

3.7.1　对称随机游动与反正弦律

设 $Y_1,Y_2,\cdots$ 独立同分布且

$$P\{Y_i=1\}=P\{Y_i=-1\}=\frac{1}{2}$$

定义

$$Z_0=0,\quad Z_n=\sum_{i=1}^{n} Y_i$$

过程 $\{Z_n,n\geqslant 0\}$ 称为对称随机游动过程。

若我们定义

$$X_n=\begin{cases}0, & 若 Z_n=0 \\ 1, & 若 Z_n>0 \\ -1, & 若 Z_n<0\end{cases}$$

则$\{X_n,n\geqslant 1\}$是一个再生过程，每当X_n取值0时过程再生。为了得出这个再生过程的一些性质，我们先研究对称随机游动$\{Z_n,n\geqslant 0\}$。

令

$$u_n=P\{Z_{2n}=0\}=\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}$$

且注意到

$$u_n=\frac{2n-1}{2n}u_{n-1} \tag{3.7.1}$$

我们回想一下，从第一章例1.5(d)（选票问题的例子）的结果可得，在对称随机游动中，在时刻$2n$首次到达0的概率表达式为

$$P\{Z_1\neq 0,Z_2\neq 0,\cdots,Z_{2n-1}\neq 0,Z_{2n}=0\}=\frac{\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}}{2n-1}=\frac{u_n}{2n-1} \tag{3.7.2}$$

我们将需要下列引理：它说明对称随机游动在时刻$2n$处于状态0的概率u_n，也等于此随机游动到时刻$2n$尚未击中0的概率。

引理 3.7.3

$$P\{Z_1\neq 0,Z_2\neq 0,\cdots,Z_{2n}\neq 0\}=u_n$$

证明　从(3.7.2)可见

$$P\{Z_1\neq 0,\cdots,Z_{2n}\neq 0\}=1-\sum_{k=1}^{n}\frac{u_k}{2k-1}$$

因此我们必须证明

$$u_n=1-\sum_{k=1}^{n}\frac{u_k}{2k-1} \tag{3.7.3}$$

我们将按n归纳进行证明。当$n=1$时，因$u_1=\frac{1}{2}$，上式成立。所以假设(3.7.3)对$n-1$成立。现在

$$\begin{aligned}1-\sum_{k=1}^{n}\frac{u_k}{2k-1}&=1-\sum_{k=1}^{n-1}\frac{u_k}{2k-1}-\frac{u_n}{2n-1}\\&=u_{n-1}-\frac{u_n}{2n-1}\qquad\text{（由归纳法假设）}\\&=u_n\qquad\text{（由(3.7.1)）}\end{aligned}$$

证毕。

因为

$$u_n=\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}$$

利用斯特林近似公式($n!\ \sim n^{n+1/2}e^{-n}\sqrt{2\pi}$)得

$$u_n\sim\frac{(2n)^{2n+1/2}e^{-2n}\sqrt{2\pi}}{n^{2n+1}e^{-2n}(2\pi)2^{2n}}=\frac{1}{\sqrt{n\pi}}$$

从而$n\to\infty$时$u_n\to 0$。于是从引理 3.7.3 可见以概率 1 对称随机游动要返回原点。

下一命 题给出直到(包括)时刻 $2n$ 为止最后一次到达 0 的时刻的分布。

命题 3.7.4

对 $k=0,1,\cdots,n$,

$$P\{Z_{2k}=0,Z_{2k+1}\neq 0,Z_{2k+2}\neq 0,\cdots,Z_{2n}\neq 0\}=u_k u_{n-k}$$

证明

$$\begin{aligned}&P\{Z_{2k}=0,Z_{2k+1}\neq 0,\cdots,Z_{2n}\neq 0\}\\&=P\{Z_{2k}=0\}P\{Z_{2k+1}\neq 0,\cdots,Z_{2n}\neq 0\mid Z_{2k}=0\}\\&=u_k u_{n-k}\end{aligned}$$

其中用了引理 3.7.3 去计算上式右端的第二项。

我们现在已作好准备讨论主要结果:如用直线连接 Z_k 与 Z_{k+1},绘出对称随机游动(起始于 $Z_0=0$,见图 3.7.1),则到时刻 $2n$ 过程有 $2k$ 个单位时间为正而 $2n-2k$ 个单位时间为负的概率与命题 3.7.4 中给出的概率相同。(对图 3.7.1 中所示的随机游动的样本路径,前 8 个单位时间中有 6 个为正,2 个为负。)

定理 3.7.5 以 $E_{k,n}$记到时刻 $2n$ 对称随机游动有 $2k$ 个单位时间为正及 $2n-2k$ 个单位时间为负这一事件,令 $b_{k,n}=P(E_{k,n})$。则

$$b_{k,n}=u_k u_{n-k} \tag{3.7.4}$$

证明 对 n 用归纳法。因为

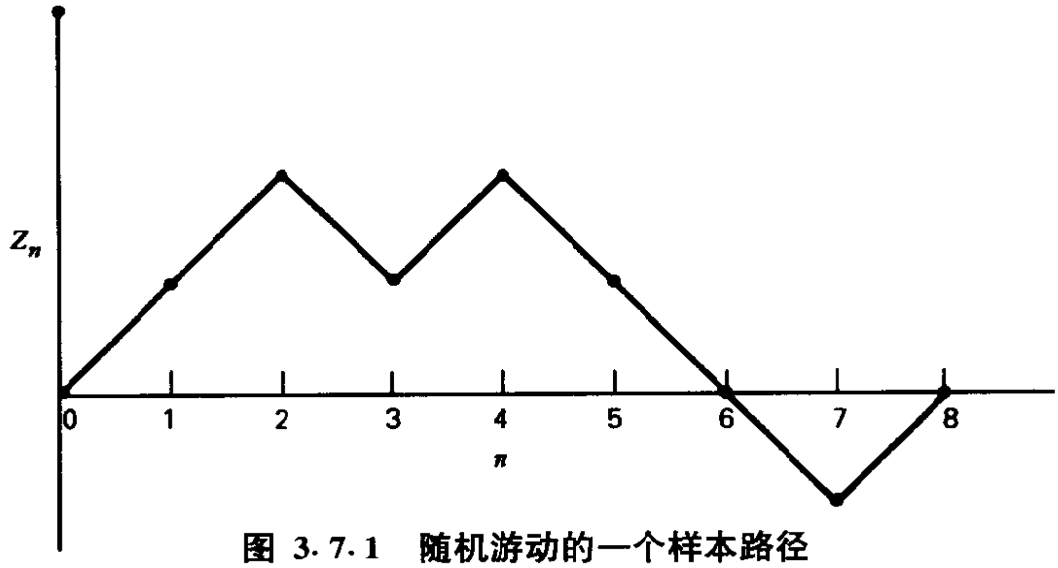


图 3.7.1 随机游动的一个样本路径

$$b_{0,1}=b_{1,1}=\frac{1}{2},\quad u_0=1,\quad u_1=\frac{1}{2}$$

得 $n=1$ 时(3.7.4)是对的。所以假定 $b_{k,m}=u_k u_{m-k}$ 对一切使得 $m<n$ 的 m 成立。为证(3.7.4)我们首先考虑 $k=n$ 的情形。对首次回到 0 的时刻 T 取条件得

$$b_{n,n}=\sum_{r=1}^{n}P\{E_{n,n}|T=2r\}P\{T=2r\}+P\{E_{n,n}|T>2n\}P\{T>2n\}$$

给定 $T=2r$，随机游动在时刻 $2r$ 首次为 0 而在$(0,2r)$中总是正或总是负是等可能的。因此，

$$P\{E_{n,n}|T=2r\}=b_{n-r,n-r}/2,\quad P\{E_{n,n}|T>2n\}=\frac{1}{2}$$

从而

$$\begin{aligned}b_{n,n}&=\frac{1}{2}\sum_{r=1}^{n}b_{n-r,n-r}P\{T=2r\}+\frac{1}{2}P\{T>2n\}\\&=\frac{1}{2}\sum_{r=1}^{n}u_{n-r}P\{T=2r\}+\frac{1}{2}P\{T>2n\}\end{aligned}$$

其中的最后一个等式 $b_{n-r,n-r}=u_{n-r}u_0$，由归纳法假设而得。现在

$$\begin{aligned}\sum_{r=1}^{n}u_{n-r}P\{T=2r\}&=\sum_{r=1}^{n}P\{Z_{2n-2r}=0\}P\{T=2r\}\\&=\sum_{r=1}^{n}P\{Z_{2n}=0|T=2r\}P\{T=2r\}\\&=u_n\end{aligned}$$

从而

$$b_{n,n}=\frac{1}{2}u_n+\frac{1}{2}P\{T>2n\}$$
$$=\frac{1}{2}u_n+\frac{1}{2}u_n \qquad (\text{据引理 } 3.7.3)$$
$$=u_n$$

因此(3.7.4)对 $k=n$ 成立，且事实上由对称性 $k=0$ 时(3.7.4)也成立。(3.7.4)对 $0<k<n$ 成立可用类似的证法。再对 T 取条件得

$$b_{k,n}=\sum_{r=1}^{n}P\{E_{k,n}|T=2r\}P\{T=2r\}$$

给定 $T=2r$，而随机游动在 $(0,2r)$ 中总是正或总是负是等可能的。因此要 $E_{k,n}$ 发生，从时刻 $2r$ 到 $2n$ 的延拓在前一情形将有 $2k-2r$ 个正单位而在后一情形需 $2k$ 个。因此

$$b_{k,n}=\frac{1}{2}\sum_{r=1}^{n}b_{k-r,n-r}P\{T=2r\}+\frac{1}{2}\sum_{r=1}^{n}b_{k,n-r}P\{T=2r\}$$
$$=\frac{1}{2}u_{n-k}\sum_{r=1}^{n}u_{k-r}P\{T=2r\}+\frac{1}{2}u_k\sum_{r=1}^{n}u_{n-r-k}P\{T=2r\}$$

其中最后的等式由归纳法假设而得。因为

$$\sum_{r=1}^{n}u_{k-r}P\{T=2r\}=u_k$$

$$\sum_{r=1}^{n}u_{n-r-k}P\{T=2r\}=u_{n-k}$$

可见

$$b_{k,n}=u_k u_{n-k}$$

证毕。

定理 3.7.5 给出的概率分布，即

$$P\{X=2k\}=u_k u_{n-k}$$

称为离散的反正弦分布。我们之所以这样称呼，是因为当 k 与 n 很大的，由斯特林近似公式有

$$u_k u_{n-k}\sim\frac{1}{\pi\sqrt{k\ (n-k)}}$$

因此对任意的 x，$0<x<1$，可见对称随机游动在 $(0,2n)$ 中为正的时间所占比例小于 x 的概率为

$$(3.7.5)\qquad \sum_{k=0}^{nx} u_k u_{n-k} \approx \frac{1}{\pi}\int_0^{nx} \frac{1}{\sqrt{y\ (n-y)}}dy$$

$$= \frac{1}{\pi}\int_0^{x} \frac{1}{\sqrt{w\ (1-w)}}dw \qquad (\text{取 } w=y/n)$$

$$= \frac{2}{\pi}\arcsin\sqrt{x}$$

从上可见，当 n 很大时，在前 $2n$ 个单位时间中对称随机游动为正的时间所占比例渐近服从由(3.7.5)给出的反正弦分布。例如，为正的时间少于一半的概率是$(2/\pi)\arcsin\sqrt{\frac{1}{2}}=\frac{1}{2}$。

上述定理的一个有趣的推论是，它告诉我们对称随机游动为正的时间所占比例不收敛于常数值1/2(因为，不然的话此极限分布就不是反正弦律而是常值随机变量的分布了)。因此如果我们考虑再生过程$\{X_n\}$，它跟踪着对称随机游动的符号，可得 X_n 等于1的时间所占比例肯定不收敛于某个常数。另一方面，由对称性及 $n\to\infty$ 时 $u_n\to 0$，显然得 $n\to\infty$ 时

$$P\{X_n = 1\} = P\{Z_n > 0\} \to \frac{1}{2}$$

上述事实(一个再生过程处于某个状态的极限概率不等于长时间后它处于该状态的时间所占的比例)不成为一个矛盾的原因是一个循环时间的期望为无限的，即对称随机游动相继到达状态0之间的平均时间

$$E[T] = \infty$$

上式必定成立，否则我们将会有矛盾。它也可直接从引理3.7.3得证，因为

$$E[T] = \sum_{n=0}^{\infty} P\{T > 2n\}$$

$$= \sum_{n=0}^{\infty} u_n \qquad (\text{由引理 } 3.7.3)$$

又由 $u_n \sim (\sqrt{n\pi})^{-1}$ 可见

$$E[T] = \infty$$

注记　从命题3.7.4及斯特林近似公式可得当$0<x<1$且n很大时有

$$
\begin{aligned}
P\{在 2nx 与 2n 之间无零点\} &= 1-\sum_{k=nx}^{n} u_k u_{n-k} \\
&= \sum_{k=0}^{nx-1} u_k u_{n-k} \\
&\approx \frac{2}{\pi}\arcsin\sqrt{x}
\end{aligned}
$$

其中最后的近似式从(3.7.5)而得。

3.8　平稳点过程

一个具有平稳增量的计数过程$\{N(t),t\geqslant 0\}$称为平稳点过程。由定理3.5.2我们注意到，平衡更新过程是平稳点过程的一个例子。

定理3.8.1　对任一平稳点过程，除对一切$t\geqslant 0$，$P\{N(t)=0\}=1$的不足道的情形外，有

(3.8.1)
$$\lim_{t\to 0}\frac{P\{N(t)>0\}}{t}=\lambda>0$$

其中不排除$\lambda=\infty$。

证明　令$f(t)=P\{N(t)>0\}$且注意到$f(t)$非负不减，又

$$
\begin{aligned}
f(s+t) &= P\{N(s+t)-N(s)>0 \quad 或 \quad N(s)>0\} \\
&\leqslant P\{N(s+t)-N(s)>0\}+P\{N(s)>0\} \\
&= f(t)+f(s)
\end{aligned}
$$

因此

$$f(t)\leqslant 2f\left(\frac{t}{2}\right)$$

且由归纳法可得，对一切$n=1,2,\cdots$

$$f(t)\leqslant nf(t/n)$$

设a使得$f(a)>0$，则对一切$n=1,2,\cdots$

(3.8.2)
$$\frac{f(a)}{a}\leqslant\frac{f(a/n)}{a/n}$$

现定义 $\lambda = \limsup_{t \to 0} f(t)/t$ 。由（3.8.2）得

$$\lambda \geqslant \frac{f(a)}{a} > 0$$

为证明 $\lambda = \lim_{t \to 0} f(t)/t$，考虑两种情形，首先假设 $\lambda < \infty$。此时固定 $\varepsilon > 0$，设 $s > 0$ 使得 $f(s)/s > \lambda - \varepsilon$。对任意的 $t \in (0,s)$ 存在一个正整数 n，使得

$$\frac{s}{n} \leqslant t < \frac{s}{n-1}$$

从 $f(t)$ 的单调性及(3.8.2)，对此区间中的一切 t 有

$$\frac{f(t)}{t} \geqslant \frac{f(s/n)}{s/(n-1)} = \frac{n-1}{n} \frac{f(s/n)}{s/n} \geqslant \frac{n-1}{n} \frac{f(s)}{s} \tag{3.8.3}$$

因此

$$\frac{f(t)}{t} > \frac{n-1}{n}(\lambda - \varepsilon)$$

既然 ε 是任意的，且 $t \to 0$ 时有 $n \to \infty$，得 $\lim_{t \to 0} f(t)/t = \lambda$。

现在假定 $\lambda = \infty$，此时固定任意大的 $A > 0$，选取 s 使得 $f(s)/s > A$。由(3.8.3) 对一切 $t \in (0,s)$ 有

$$\frac{f(t)}{t} \geqslant \frac{n-1}{n} \frac{f(s)}{s} > \frac{n-1}{n} A$$

此式蕴含 $\lim_{t \to 0} f(t)/t = \infty$，证毕。

例 3.8(a) 对于平衡更新过程，

$$P\{N(t) > 0\} = F_e(t) = \int_0^t \bar{F}(y)dy/\mu$$

因此，用洛比达法则得

$$\lambda = \lim_{t \to 0} \frac{P\{N(t) > 0\}}{t} = \lim_{t \to 0} \frac{\bar{F}(t)}{\mu} = \frac{1}{\mu}$$

所以，对于平衡更新过程，λ 是更新过程的速率。

对任一平稳点过程 $\{N(t), t \geqslant 0\}$，有

$$E[N(t+s)] = E[N(t+s) - N(s)] + E[N(s)] = E[N(t)] + E[N(s)]$$

故有某个常数 c 使

$$E[N(t)] = ct$$

c 与 λ 之间有什么关系呢?（在平衡更新过程的情形，从例 3.8(a)

及定理 3.5.2 得 $\lambda=c=1/\mu$。）一般地，我们注意到

$$c=\sum_{n=1}^{\infty}\frac{nP\{N(t)=n\}}{t}$$
$$\geqslant\sum_{n=1}^{\infty}\frac{P\{N(t)=n\}}{t}$$
$$=\frac{P\{N(t)>0\}}{t}$$

所以得 $c\geqslant\lambda$。为确定何时 $c=\lambda$，我们需要下述概念，平稳点过程称为正则的或有序的，若

(3.8.4) $$P\{N(t)\geqslant 2\}=o(t)$$

应注意到，对于一平稳点过程(3.8.4)蕴含着在任意时刻同时发生两个或多个事件的概率是 0。为了明了这点，把区间[0,1]分为 n 个相等的部份，事件同时发生的概率小于在

$$\left(\frac{j}{n},\frac{j+1}{n}\right),\quad j=0,1,2,\cdots,n-1$$

中的任一区间中发生两个或更多的事件的概率，后者小于 $nP\{N(1/n)\geqslant 2\}$，而由(3.8.4)当 $n\to\infty$ 时它趋于 0。当 $c<\infty$，逆命题亦真，即任何平稳点过程，若 c 有限且多个事件同时发生的概率是零，则必是正则的，其证明更为复杂，不再给出。

将对正则平稳点过程证明 $c=\lambda$ 以结束本节。此即柯罗留克定理。

柯罗留克定理 对于一正则平稳点过程，每单位时间事件发生的平均数 c 与(3.8.1)定义的强度 λ 相等。不排除 $\lambda=c=\infty$ 的情形。

证明 定义下列记号：

A_k 为事件 $\{N(1)>k\}$；

B_{nj} 为事件 $\left\{N\left(\frac{j+1}{n}\right)-N\left(\frac{j}{n}\right)\geqslant 2\right\}$；

$B_n=\bigcup_{j=0}^{n-1}B_{nj}$；

C_{nkj} 为事件 $\left\{N\left(\frac{j+1}{n}\right)-N\left(\frac{j}{n}\right)\geqslant 1, N(1)-N\left(\frac{j+1}{n}\right)=k\right\}$

设 $\epsilon>0$ 及正整数 m 给定，从所假设的过程的正则性可知对一切充分大的 n

$$P(B_{nj})<\frac{\epsilon}{n(m+1)},\qquad j=0,1,\cdots,n-1$$

因此

$$P(B_n)\leqslant\sum_{j=0}^{n-1}P(B_{nj})\leqslant\frac{\epsilon}{m+1}$$

所以

(3.8.5)
$$\begin{aligned}P(A_k)&=P(A_k\bar{B}_n)+P(A_kB_n)\\&\leqslant P(A_k\bar{B}_n)+\frac{\epsilon}{m+1}\end{aligned}$$

其中 $\bar{B}_n$ 是 B_n 的余集，然而稍加思索可得

$$A_k\bar{B}_n=\bigcup_{j=0}^{n-1}C_{nkj}\bar{B}_n$$

因此

$$P(A_k\bar{B}_n)\leqslant\sum_{j=0}^{n-1}P(C_{nkj})$$

这连同(3.8.5)蕴含了对一切充分大的 n 有

(3.8.6
$$\begin{aligned}\sum_{k=0}^{m}P(A_k)&\leqslant\sum_{j=0}^{n-1}\sum_{k=0}^{m}P(C_{nkj})+\epsilon\\&=\sum_{j=0}^{n-1}P\left\{N\left(\frac{j+1}{n}\right)-N\left(\frac{j}{n}\right)\geqslant 1,\right.\\&\qquad\left.N(1)-N\left(\frac{j+1}{n}\right)\leqslant m\right\}+\epsilon\\&\leqslant\sum_{j=0}^{n-1}P\left\{N\left(\frac{j+1}{n}\right)-N\left(\frac{j}{n}\right)\geqslant 1\right\}+\epsilon\\&=nP\left\{N\left(\frac{1}{n}\right)\geqslant 1\right\}+\epsilon\\&\leqslant\lambda+2\epsilon\end{aligned}$$

因(3.8.6)对一切 m 成立，故得

$$\sum_{k=0}^{\infty}P(A_k)\leqslant\lambda+2\epsilon$$

因此

$$c=E[N(1)]=\sum_{k=0}^{\infty}P\{N(1)>k\}=\sum_{k=0}^{\infty}P(A_k)\leqslant\lambda+2\epsilon$$

由于 ϵ 是任意的且早已知道 $c\geqslant\lambda$，结论得证。

习 题

1. 下列命题是否正确：

   (a) $N(t)<n$ 当且仅当 $S_n>t$?

   (b) $N(t)\leqslant n$ 当且仅当 $S_n\geqslant t$?

   (c) $N(t)>n$ 当且仅当 $S_n<t$?

2. 在定义一个更新过程中我们假设来到间隔为有限的概率 $F(\infty)$等于 1。若 $F(\infty)<1$，那么在每次更新之后以正概率 $1-F(\infty)$不再有别的更新。证明：当 $F(\infty)<1$ 时更新的总数，记为 $N(\infty)$，使得 $1+N(\infty)$有几何分布，均值为 $1/(1-F(\infty))$。

3. 用语言表述随机变量 $X_{N(t)+1}$代表什么。(提示：它是哪个更新区间的长度?)证明

$$P\{X_{N(t)+1}\geqslant x\}\geqslant \bar{F}(x)$$

当 $F(x)=1-e^{-\lambda x}$时，精确地计算上式的左端。

4. 证明更新方程

$$m(t)=F(t)+\int_0^t m(t-x)dF(x)$$

5. 随机变量 X_1，…，X_n 称为可交换的，若 i_1，i_2，…，i_n 是 1，2，…，n 的一个置换，则 X_{i_1}，…，X_{i_n}与 X_1，…，X_n 有相同的联合分布，也就是，若 $P\{X_1\leqslant x_1,X_2\leqslant x_2,\cdots,X_n\leqslant x_n\}$ 是 $(x_1,x_2,\cdots x_n)$ 的一个对称函数，则它们是可交换的。以 X_1，X_2，…记一个更新过程的来到间隔。

   (a) 证明：在 $N(t)=n$ 的条件之下 X_1，…X_n 是可交换的。又 X_1，…，X_n，X_{n+1}是可交换的吗(在 $N(t)=n$ 的条件之下)?

   (b) 用(a)证明：对 $n>0$

$$E\left[\frac{X_1+\cdots+X_{N(t)}}{N(t)}\,\middle|\,N(t)=n\right]=E[X_1\mid N(t)=n]$$

   (c) 证明：

$$E\left[\frac{X_1+\cdots+X_{N(t)}}{N(t)}\,\middle|\,N(t)>0\right]=E[X_1\mid X_1<t]$$

6. 考虑一个单服务员的银行，可能来的顾客按参数为 λ 的泊松过程来到，但仅在他们来到时服务员有空的情况下才进入银行。以 G 记服务时间分布。

(a) 顾客进入银行的速率是多少？

(b) 多少比例的潜在的顾客进入银行？

(c) 服务员的忙期占时间的多少比例？

7. 设 X_1，X_2，…独立同分布，$E[X_i]<\infty$。又设 N_1，N_2，…独立同分布且是序列 X_1，X_2，…的停时，$E[N_i]<\infty$。序贯地观察 X_i，在 N_1 停止。然后在余下的 X_i 中取样(做法好像取样恰从 X_{N_1+1}开始一样)，再观察 N_2 个后停止。(于是例如，$X_1+\cdots+X_{N_1}$与 $X_{N_1+1}+\cdots+X_{N_1+N_2}$同分布)。现在再在剩下的 X_i 中取样（做法又好像取样刚开始一样）再观察 N_3 个后停止，等等。

(a) 令

$$S_1=\sum_{i=1}^{N_1}X_i,\quad S_2=\sum_{i=N_1+1}^{N_1+N_2}X_i,\quad \cdots,\quad S_m=\sum_{i=N_1+\cdots+N_{m-1}+1}^{N_1+\cdots+N_m}X_i$$

用强大数定律计算

$$\lim_{m\to\infty}\left(\frac{S_1+\cdots+S_m}{N_1+\cdots+N_m}\right)$$

(b) 记

$$\frac{S_1+\cdots+S_m}{N_1+\cdots+N_m}=\frac{S_1+\cdots+S_m}{m}\cdot\frac{m}{N_1+\cdots+N_m}$$

对(a)中的极限导出另一表达式。

(c) 将两种表达式列成等式而得瓦尔德等式。

8. 一位矿工困于一矿井中，此矿井有三道门。1 号门引导他经两天的行程获救脱险；2 号门引导他经四天的行程又回到这矿井中；3 号门领他经八天的行程又回到这矿井中。假设在任何时刻他等可能地选择这三道门之一，且以 T 记矿工获救脱险所用的时间。

(a) 定义一列独立同分布随机变量 X_1，X_2，…及一停时 N 使得

$$T=\sum_{i=1}^{N}X_i$$

注意：你可以设想即使到达安全地后矿工仍然继续随机地挑选门。

(b) 用瓦尔德等式求 $E[T]$。

(c) 计算 $E\left[\sum_{i=1}^{n}X_i\mid N=n\right]$且注意到它不等于 $E\left[\sum_{i=1}^{n}X_i\right]$。

(d) 利用(c)再次推导出 $E[T]$。

9. 表明如何从关键更新定理推出布莱克威尔定理。

10. 一个过程处在 n 个状态 1，2，…，n 之一。最初它处于状态 1，停留在该状态的时间有分布 F_1。离开状态 1 后它到状态 2，在那儿停留的时间有分布 F_2。离开 2 后它到状态 3，等等。从状态 n 它回到 1，且再从头开始，求：

$$\lim_{t\to\infty} P\{\text{在时刻 } t \text{ 过程处于状态 } i\}$$

假设进到状态 1 之间的时间的分布 H 是非格点的且均值有限。

11. 以 $A(t)$ 与 $Y(t)$ 记一个更新过程在时刻 t 的年龄与剩余寿命。填空：

(a) $A(t) > x \Leftrightarrow$ 在______区间中没发生事件?

(b) $Y(t) > x \Leftrightarrow$ 在______区间中没发生事件?

(c) $P\{Y(t) > x\} = P\{A(\quad) > \quad\}$。

(d) 对泊松过程计算 $A(t)$ 与 $Y(t)$ 的联合分布。

12. 以 $A(t)$ 与 $Y(t)$ 分别记在时刻 t 的年龄与剩余寿命。求：

(a) $P\{Y(t) > x \mid A(t) = s\}$。

(b) $P\left\{Y(t) > x \mid A\left(t + \frac{x}{2}\right) = s\right\}$。

(c) 对泊松过程求 $P\{Y(t) > x \mid A(t+x) > s\}$。

(d) $P\{Y(t) > x, A(t) > y\}$。

(e) 若 $\mu < \infty$，证明：以概率 1，$t\to\infty$ 时 $A(t)/t \to 0$。

13. 考虑一个更新过程，其来到间隔分布是参数为 (n,λ) 的 Γ-分布。利用命题 3.4.6 证明：

$$\lim_{t\to\infty} E[Y(t)] = \frac{n+1}{2\lambda}$$

解释为何可以不作任何计算而得上式。

14. 一个形为

$$g(t) = h(t) + \int_0^t g(t-x)dF(x)$$

的方程称为更新型方程。用卷积记号可记为

$$g = h + g * F$$

将上式叠代或用拉普拉斯变换证明：更新型方程有解

$$g(t) = h(t) + \int_0^t h(t-x)dm(x)$$

其中 $m(x) = \sum_{n=1}^{\infty} F_n(x)$。若 h 直接黎曼可积，且 F 非格点、均值有限，

则运用关键更新定理可得

$$\lim_{t\to\infty} g(t) = \frac{\int_0^\infty h(t)dt}{\int_0^\infty \bar{F}(t)dt}$$

对过程在概率上重新开始的时刻取条件可得 $g(t)$ 的更新型方程。对以下函数建立更新型方程。

(a) $P(t)$，一个交错更新过程在时刻 t 是开着的概率。

(b) $g(t) = E[A(t)]$，一个更新过程在时刻 t 的年龄的期望。

应用关键更新定理求(a)与(b)中的极限值。

15. 在习题 6 中假设顾客按更新过程来到，来到间隔分布为 F。若一个事件对应于一个

(a) 走进银行的顾客，

(b) 离开银行的顾客，

到时刻 t 事件的个数将构成一个(可能是延迟的)更新过程吗？若 F 是指数分布结果怎样？

16. 证明方程(3.5.3)。

17. 考虑接连地投掷一均匀硬币，以 H 与 T 分别记正面与反面。

(a) 计算直到花样 HHTHHTT 出现的平均投掷次数。

(b) HHTT 或 HTHT，哪一个花样出现的平均时间更长？

18. 考虑一个延迟更新过程 $\{N_D(t), t \geqslant 0\}$，其第一个来到间隔分布为 G，而其它来到间隔分布为 F。令 $m_D(t) = E[N_D(t)]$。

(a) 证明：

$$m_D(t) = G(t) + \int_0^t m(t-x)dG(x)$$

其中 $m(t) = \sum_{n=1}^{\infty} F_n(t)$ 。

(b) 以 $A_D(t)$ 记时刻 t 的年龄。证明：若 F 不是格点的，$\int x^2 dF(x) < \infty$ 且 $t\to\infty$ 时 $t\bar{G}(t)\to 0$，则

$$E[A_D(t)] \to \frac{\int_0^\infty x^2 dF(x)}{2\int_0^\infty x dF(x)}$$

(c) 证明：若 G 有有限均值，则 $t\to\infty$ 时 $t\bar{G}(t)\to 0$。

19. 对更新酬劳过程证明布莱克威尔定理，也就是，假设循环分布不是格点的，证明 $t\to\infty$ 时

$$E[\text{在}(t,t+a)\text{ 中的酬劳}]\to a\,\frac{E[\text{一个循环中的酬劳}]}{E[\text{一个循环的时间}]}$$

假定每个有关的函数都是直接黎曼可积的。

20. 对更新酬劳过程证明

$$\lim_{t\to\infty}E[R_{N(t)+1}]=\frac{E[R_1X_1]}{E[X_1]}$$

假定 X_i 的分布不是格点的，而且每个有关的函数都是直接黎曼可积的。当循环的酬劳定义为与循环的长度相等时，上式导至

$$\lim_{t\to\infty}E[X_{N(t)+1}]=\frac{E[X^2]}{E[X]}$$

除非 X 以概率 1 是常数，上式总是大于 $E[X]$（为什么？）。

21. 在例 3.6(c)中假设来到的更新过程是泊松过程，来到间隔的均值为 μ。每当有 N 个乘客等待时就开出一列火车，以 N^* 记使长时间后的平均费用最小的 N 值。另一类决策是每 T 个单位时间开出一列火车。计算在这种决策下的长时间后的平均费用，且以 T^* 记使它最小的 T 值。证明：每当有 N^* 个顾客等待时开出火车的平均费用小于每 T^* 个单位时间开出火车的平均费用。

22. 一辆小汽车的寿命是分布为 F 的随机变量。当汽车损坏或用了 A 年时，车主就以旧换新。以 $R(A)$ 记一辆用了 A 年的旧车的卖出价格。一辆损坏的车没有任何价值。以 C_1 记一辆新车的价格，且假设每当车子损坏时还要额外承担费用 C_2。

(a) 每当购置一新车时就说一个循环开始，计算长时间后的单位时间的平均费用。

(b) 每当使用中的汽车损坏时就说一个循环开始，计算长时间后的单位时间的平均费用。

注意：不仅在(a)中而且在(b)中都要你计算一个循环中的平均费用与一个循环的平均时间之比。当然，(a)与(b)的答案应当是同样的。

23. 考虑一个单服务员的排队系统，具有参数为 λ 的泊松来到而服务时间分布 G 具有均值 μ_G。假设 $\lambda\mu_G<1$。

(a) 求系统是空的时间所占比例 P_0。

(b) 当系统不空时就说系统正忙着（所以服务员正忙着），计算一个忙期

的平均长度。

(c)利用(b)及瓦尔德等式计算在一个忙期中受到服务的顾客的平均数。

24. 对于 3.6.1 节的排队系统，定义 $V(t)$，时刻 t 系统的工作量，为在时刻 t 系统中的全部顾客所剩余的服务时间之和。令

$$V = \lim_{t\to\infty}\int_0^t V(s)ds/t$$

又以 D_i 记第 i 个顾客排队等待所用去的时间，定义

$$W_Q = \lim_{n\to\infty}(D_1 + \cdots + D_n)/n$$

(a) 论证 V 与 W_Q 存在，且以概率 1 为常数。

(b) 证明恒等式

$$V = \lambda E[Y]W_Q + \lambda E[Y^2]/2$$

其中 $1/\lambda$ 是平均来到间隔时间而 Y 具有服务时间的分布。

25. 举反例证明，在 k 个服务员且来到为更新过程的排队模型中，条件 $E[Y]<kE[X]$ 并非是循环时间必为有限的充分条件，其中 Y 是服务时间而 X 是来到间隔。(提示：给出一例，在第一个来到之后系统永远不空。)

26. 包裹依照参数为 λ 的泊松过程来到一邮寄点。收取所有待运的包裹的卡车的来到遵循一个具有非格点的来到间隔分布 F 的更新过程。以$X(t)$记在时刻 t 等待装车的包裹个数。

(a) $\{X(t), t\geqslant 0\}$ 是什么类型的随机过程?

(b) 求：

$$\lim_{t\to\infty}P\{X(t)=i\}, \qquad i\geqslant 0$$

的表达式。

27. 考虑一个满足定理 3.7.1 的条件再生过程。假设每当过程处于状态 j 就以比率 $r(j)$ 得到报酬。若一个循环中的平均酬劳是有限的，证明：长时间后单位时间的平均酬劳以概率 1 为

$$\lim_{t\to\infty}\int_0^t \frac{r(X(s))ds}{t} = \sum_j P_j r(j)$$

其中 P_j 是 $X(t)$ 等于 j 的极限概率。

参考文献

文献 9 提供了一篇有启发性的更新理论的评论文章，关于在最严格的条

件下的关键更新定理的证明，读者应当参考费勒的第二卷(文献 4)，定理 3.6.2 在排队论文献中以“里特尔公式”著称。我们处理反正弦律的方法类似于费勒第一卷(文献 4)中的方法，有兴趣的读者也应当参看文献 3。

1. D. R. Cox, *Renewal Theory*, Methuen, London, 1962。
2. H. Cramer and M. Leadbetter, *Stationary and Related Stochastic Proce-sses*, Wiley, New York, 1966。
3. De Finetti, *Theory of Probability*, Vol. 1, Wiley, New York, 1970。
4. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. I. and Ⅱ, Wiley, New York, 1957 and 1966。
5. S. Karlin and H. Taylor, *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York, 1975。
6. S. Ross, *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco, 1970。
7. S. Ross, Introduction to Probability Models, 2nd ed., Academic Press, New York, 1981。
8. W. Smith, "Regenerative Stochastic Processes," in *Proceedings Royal Society*, Series A, 232 (1955), pp. 6—31。
9. W. Smith, "Renewal Theory and Its Ramifications," *Journal of the Royal* Statistical Society, Series B, 20 (1958), pp. 243-302。

文献 1,5 及 6 以与本教材大致相同的数学水平介绍更新理论。文献 7 给出了一个更简单更直观的方法。

4 马尔可夫链

4.1 引言与例子

考虑只取有限个或可数个值的随机过程$\{X_n, n=0,1,2\cdots\}$。若不另外说明，过程可能取得值的集合将以非负整数集$\{0,1,2,\cdots\}$来表示。若$X_n=i$，就说过程在时刻n处于状态i，假设每当过程处于状态i，则在下一时刻将处于状态j的概率是固定的P_{ij}。也即假设对一切状态$i_0,i_1,\cdots,i_{n-1},i,j$及一切$n\geqslant 0$有

(4.1.1)
$$P\{X_{n+1}=j \mid X_n=i, X_{n-1}=i_{n-1},\cdots,X_1=i_1,X_0=i_0\} = P_{ij}$$

这样的随机过程称为马尔可夫链。方程(4.1.1)可解释为，对马尔可夫链，给定过去的状态$X_0,X_1,\cdots,X_{n-1}$及现在的状态X_n，将来的状态X_{n+1}的条件分布与过去的状态独立，只依赖于现在的状态，这称为马尔可夫性。P_{ij}代表处于状态i的过程下一步转移到状态j的概率。由于概率是非负的，且过程必须转移到某个状态，所以有

$$P_{ij}\geqslant 0,\quad i,j\geqslant 0;\quad \sum_{j=0}^{\infty} P_{ij}=1,\quad i=0,1,\cdots$$

以P记一步转移概率P_{ij}的矩阵，从而

$$P=\begin{pmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & & & \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

例 4.1（a） $M/G/1$ 排队系统。 假设顾客依照参数为λ的泊松过程来到一服务中心，只有一个服务员，来客发现服务员空着即刻得到服务；其他人排队等待直至轮到他们。相继的顾客的服务时间假定是独立的随机变量，具有共同的分布G；且也假定他们与来到过程独立。

上述系统称为$M/G/1$排队系统。字母M代表顾客来到间隔的分布是指数分布，G代表服务时间的分布；数字1表示只有一个服务员。

若以$X(t)$记在时刻t系统中的顾客人数，则$\{X(t),t\geq 0\}$不具有马尔可夫性（即将来状态的条件分布只依赖于现在而不依赖于过去）。因为若我们知道在时刻t系统中的人数，那么为了预测未来的状态，我们不用关心从最近的一位顾客来到后已过去了多少时间（因为来到过程是无记忆的），但要注意在服务中的顾客已服务了多长时间（因为服务分布G是任意的，故不是无记忆的）。

克服所述困难的一种方法是我们只在顾客离去的时刻考察系统。即以X_n记第n个顾客走后留下的顾客数，$n\geq 1$。又以Y_n记第$(n+1)$个顾客受服务期间来到的顾客数。

当$X_n>0$时，第n个顾客离去后余下有X_n个顾客，其中一人进入服务，而其余X_n-1人排队等候。因此，下一个人离去时系统中将包含在排队的X_n-1个顾客以及第$n+1$个顾客服务期间的来客。当$X_n=0$时可作类似的讨论，由此可见

(4.1.2)
$$X_{n+1}=\begin{cases} X_n-1+Y_n, & \text{若 } X_n>0 \\ Y_n, & \text{若 } X_n=0 \end{cases}$$

由于Y_n，$n\geq 1$，表示在不相重叠的服务时间区间中来到的人数，来到过程又是泊松过程，所以它们相互独立，且：

(4.1.3)
$$P\{Y_n=j\}=\int_0^\infty e^{-\lambda x}\frac{(\lambda x)^j}{j!}dG(x), \quad j=0,1,2,\cdots$$

从(4.1.2)与(4.1.3)得$\{X_n,n=1,2,\cdots\}$是马尔可夫链，转移概率为：

$$P_{0j}=\int_0^\infty e^{-\lambda x}\frac{(\lambda x)^j}{j!}dG(x), \qquad j\geq 0$$

$$P_{ij}=\int_0^\infty e^{-\lambda x}\frac{(\lambda x)^{j-i+1}}{(j-i+1)!}dG(x),\qquad j\geqslant i-1,\quad i\geqslant 1$$

$$P_{ij}=0,\qquad \text{其它}$$

例 4.1(b) G/M/1 排队系统。 假设顾客依照一个任意的更新过程来到一个单服务员的服务中心，来到间隔分布为 G。进一步假设服务分布是指数分布，参数为 μ。

若以 X_n 记第 n 个顾客来到时见到系统中的顾客数，易知过程 $\{X_n, n\geqslant 1\}$ 是马尔可夫链。为了对此马尔可夫链计算转移概率 P_{ij}，我们首先注意到，只要有顾客在接受服务，则在任意长为 t 的时间中服务完的顾客数是均值为 μt 的泊松随机变量。这是因为相继的服务时间服从指数分布，且如我们所知，这意味着服务完的顾客数构成一个泊松过程。所以，

$$P_{i,i+1-j}=\int_0^\infty e^{-\mu t}\frac{(\mu t)^j}{j!}dG(t),\quad j=0,1,\cdots,i$$

这是因为若一个来客发现有 i 个人在系统中，那么下一个来客将发现人数为 $i+1$ 减去已服务完毕的人数，易知有 j 个人被服务完毕的概率（对相继来到之间的时间取条件）等于上式的右端。

P_{i0} 的公式略有不同（它是在一个长度分布为 G 的随机时间区间内至少有计 1 个泊松事件发生的概率），由下式给出：

$$P_{i,0}=\int_0^\infty\sum_{k=i+1}^\infty e^{-\mu t}\frac{(\mu t)^k}{k!}dG(t),\quad i\geqslant 0$$

注记 读者应当注意，在前面两个例子中，我们能够发现一个嵌入马尔可夫链是因为我们在某些便于利用指数分布的无记忆性的时刻观察过程。对于出现指数分布的过程这常常是一种富有成效的方法。

例 4.1(c) 独立同分布随机变量之和。一般的随机游动。 设 X_i，$i\geqslant 1$，独立同分布，且

$$P\{X_i=j\}=a_j,\qquad j=0,\pm 1,\cdots$$

若令

$$S_0=0\quad 及\quad S_n=\sum_{i=1}^n X_i$$

则 $\{S_n, n\geqslant 0\}$ 是马尔可夫链，其转移概率

$$P_{ij}=a_{j-i}$$

$\{S_n, n\geqslant 0\}$ 称为一般随机游动，将在第七章中研究。

例 4.1(d) 简单随机游动的绝对值。随机游动$\{S_n, n\geqslant 1\}$，其中$S_n=\sum_{i=1}^{n}X_i$，称为简单随机游动，若对于某个p，$0<p<1$，有

$$P\{X_i=1\}=p$$
$$P\{X_i=-1\}=q\equiv 1-p$$

于是在简单随机游动中过程总是(以概率p)朝上一步或(以概率q)朝下一步。

现在考虑简单随机游动的绝对值$|S_n|$。过程$\{|S_n|, n\geqslant 1\}$度量了此简单随机游动在各个时刻与原点的绝对距离。有点令人惊奇的是$\{|S_n|\}$本身是一马尔可夫链。为了证明这点，我们首先证明：若$|S_n|=i$，则无论以前的值是怎么样的，S_n等于i(与$-i$相对立)的概率是$p^i/(p^i+q^i)$。

命题 4.1.1

若$\{S_n, n\geqslant 1\}$是一简单随机游动，则

$$P\{S_n=i \mid |S_n|=i, |S_{n-1}|=i_{n-1}, \cdots, |S_1|=i_1\}=\frac{p^i}{p^i+q^i}$$

证明 若令$i_0=0$，$S_0=0$，且定义

$$j=\max\{k: 0\leqslant k\leqslant n, i_k=0\}$$

因为我们知道S_j的实际值，所以显然有

$$\begin{aligned}
&P\{S_n=i \mid |S_n|=i, |S_{n-1}|=i_{n-1}, \cdots, |S_1|=i_1\}\\
&\quad=P\{S_n=i \mid |S_n|=i, \cdots, |S_{j+1}|=i_{j+1}, |S_j|=0\}
\end{aligned}$$

为使$|S_{j+1}|=i_{j+1}, \cdots, |S_n|=i$，序列$S_{j+1}, \cdots, S_n$的值只有两种可能。第一种可能导致$S_n=i$，其概率为

$$p^{\frac{n-j}{2}+\frac{i}{2}}q^{\frac{n-j}{2}-\frac{i}{2}}$$

第二种可能导致$S_n=-i$，其概率为

$$p^{\frac{n-j}{2}-\frac{i}{2}}q^{\frac{n-j}{2}+\frac{i}{2}}$$

因此，

$$\begin{aligned}
P\{S_n=i \mid |S_n|=i, \cdots, |S_1|=i_1\}&=\frac{p^{\frac{n-j}{2}+\frac{i}{2}}q^{\frac{n-j}{2}-\frac{i}{2}}}{p^{\frac{n-j}{2}+\frac{i}{2}}q^{\frac{n-j}{2}-\frac{i}{2}}+p^{\frac{n-j}{2}-\frac{i}{2}}q^{\frac{n-j}{2}+\frac{i}{2}}}\\
&=\frac{p^i}{p^i+q^i}
\end{aligned}$$

命题得证。

从命题4.1.1，并对 $S_n=+i$ 或 $-i$ 取条件得

$$P\{|S_{n+1}|=i+1 \mid |S_n|=i, |S_{n-1}|, \cdots, |S_1|\}$$

$$=P\{S_{n+1}=i+1 \mid S_n=i\}\frac{p^i}{p^i+q^i}$$

$$+P\{S_{n+1}=-(i+1) \mid S_n=-i\}\frac{q^i}{p^i+q^i}=\frac{p^{i+1}+q^{i+1}}{p^i+q^i}$$

因此，$\{|S_n|, n\geqslant 1\}$ 是一马尔可夫链，具有转移概率

$$P_{i,i+1}=\frac{p^{i+1}+q^{i+1}}{p^i+q^i}=1-P_{i,i-1}, \qquad i>0$$

$$P_{01}=1$$

4.2 切普曼——柯尔莫哥洛夫方程及状态分类

我们已经定义了一步转移概率 P_{ij}。现在我们定义 n 步转移概率 P_{ij}^n 为处于状态 i 的过程经 n 次转移后处于状态 j 的概率。即

$$P_{ij}^n=P\{X_{n+m}=j \mid X_m=i\}, \quad n\geqslant 0, \quad i,j\geqslant 0$$

当然有 $P_{ij}^1=P_{ij}$。切普曼一柯尔莫哥洛夫方程提供了计算 n 步转移概率的方法。这些方程是：对一切 n，$m\geqslant 0$，一切 i，j，有

(4.2.1) $$P_{ij}^{n+m}=\sum_{k=0}^{\infty}P_{ik}^n P_{kj}^m$$

可如下建立它们：

$$P_{ij}^{n+m}=P\{X_{n+m}=j \mid X_0=i\}$$

$$=\sum_{k=0}^{\infty}P\{X_{n+m}=j, X_n=k \mid X_0=i\}$$

$$=\sum_{k=0}^{\infty}P\{X_{n+m}=j \mid X_n=k, X_0=i\}P\{X_n=k \mid X_0=i\}$$

$$=\sum_{k=0}^{\infty}P_{kj}^n P_{ik}^n$$

若以 $P^{(n)}$ 记 n 步转移概率 P_{ij}^n 的矩阵，则由(4.2.1)断言

$$P^{(n+m)}=P^{(n)}\cdot P^{(m)}$$

其中的点代表矩阵乘法。因此，

$$P^{(n)}=P\cdot P^{(n-1)}=P\cdot P\cdot P^{(n-2)}=\cdots=P^n$$

于是$P^{(n)}$可以由矩阵P自乘n次而算出。

若对某个$n\geqslant0$有$P_{ij}^n>0$，则称从状态i可到达状态j。两个相互可到达的状态i与j称为相通的，记作$i\leftrightarrow j$。

命题 4.2.1

相通是一种等价关系，即：

(i) $i\leftrightarrow i$;

(ii) 若$i\leftrightarrow j$，则$j\leftrightarrow i$;

(iii) 若$i\leftrightarrow j$及$j\leftrightarrow k$，则$i\leftrightarrow k$。

证明 前两点从相通的定义显然得知。为了证明(iii)，假设$i\leftrightarrow j$及$j\leftrightarrow k$；则存在m,n，使得$P_{ij}^m>0,P_{jk}^n>0$。因此，

$$P_{ik}^{m+n}=\sum_{r=0}^{\infty}P_{ir}^mP_{rk}^n\geqslant P_{ij}^mP_{jk}^n>0$$

类似地可以证明存在一个s使$P_{ki}^s>0$。

称相通的两个状态属于同一个类；由命题 4.2.1，任意两个类或不相交或相同。称马尔可夫链是不可约的，若只存在一个类——即一切状态彼此相通。

称状态i有周期d，若对每个不可被d整除的n有$P_{ii}^n=0$，且d是具有此性质的最大整数。(若对一切$n>0$，$P_{ii}^n=0$，则定义i的周期是无穷大。) 具有周期 1 的状态称为非周期的。以$d(i)$记i的周期。现在我们来证明周期性是一个类性质。

命题 4.2.2 若$i\leftrightarrow j$，则$d(i)=d(j)$。

证明 设m与n使得$P_{ij}^mP_{ji}^n>0$，且假设$P_{ii}^s>0$。则：

$$P_{jj}^{n+m}\geqslant P_{ji}^nP_{ij}^m>0$$

$$P_{jj}^{n+s+m}\geqslant P_{ji}^nP_{ii}^sP_{ij}^m>0$$

例如，第二个不等式成立是因为其左端表示链开始时处于状态j，经过$n+s+m$步转移后回到j的概率，而右端是同一事件在经n步及$n+s$步后都到达i的限制下的概率。因此$d(j)$同时整除$n+m$与$n+m+s$；于是只要$P_{ii}^s>0$，$n+s+m-(n+m)=s$被$d(j)$整除。所以$d(j)$整除$d(i)$。类似论证$d(i)$整除$d(j)$。于是$d(i)=d(j)$。

对于任何状态 i 与 j，定义 f_{ij}^n 是从 i 出发在时刻 n 首次转移到 j 的概率。正式地，

$$f_{ij}^0=0$$
$$f_{ij}^n=P\{X_n=j,\ X_k\neq j,\ k=1,\ \cdots,\ n-1\mid X_0=i\}$$

令

$$f_{ij}=\sum_{n=1}^{\infty}f_{ij}^n$$

则 f_{ij} 表示过程从 i 出发迟早转移到状态 j 的概率。(注意，对 $i\neq j$，f_{ij} 为正的当且仅当从 i 可达到 j)。若 $f_{jj}=1$，称状态 j 是常返的，否则称为滑过的或非常返的。

命题 4.2.3

状态 j 为常返的当且仅当 $\sum_{n=1}^{\infty}P_{jj}^n=\infty$。

证明 若过程以概率 1 从 j 出发最终要回到 j，则状态 j 是常返的，然而由马尔可夫性可知：回到状态 j 在概率上过程就重新开始。因此，概率 1 它将再回到 j。重复这种论证可见到达状态 j 的次数以概率 1 是无限的，因而其期望是无限的。另一方面，假设 j 是滑过的，则每当过程回到 j，将有一个正概率 $1-f_{jj}$，它永远不再回到 j；因此到达 j 的次数是具有有限均值 $1/(1-f_{jj})$ 的几何分布的变量。

由上述论证可见状态 j 为常返的当且仅当

$$E\ [\text{到达 } j \text{ 的次数}\mid X_0=j]=\infty$$

但若令

$$I_n=\begin{cases}1, & \text{若 } X_n=j\\ 0, & \text{若 } X_n\neq j\end{cases}$$

$\sum_{n=0}^{\infty}I_n$ 表示到达 j 的次数。因为

$$E\Big[\sum_{n=0}^{\infty}I_n\mid X_0=j\Big]=\sum_{n=0}^{\infty}E[I_n\mid X_0=j]=\sum_{n=0}^{\infty}P_{jj}^n$$

结论得证。

导致上述命题的论证极为重要，因为它也证明了一个滑过状态只能到达有限次（由此得名滑过）。这也导出下述结论：有限状

态的马尔可夫链不可能全部状态都是滑过的。为了明瞭这点，假设各状态是 0，1，…，M，且假设它们都是滑过的，那么在一段有限时间之后（比如 T_0 时间之后）将不再进入状态 0，而在一段时间（比如 T_1）之后将不再进入状态 1，也在一段时间（比如 T_2）之后将不再进入状态 2，如此等等。于是在一段有限时间 $T=\max\{T_0, T_1,\cdots,T_M\}$ 之后将不再进入任何状态。但过程在时间 T 之后必须处于某个状态，得到矛盾。这就证明了至少有一状态必须是常返的。

我们将利用命题 4.2.3 证明常返性像周期性一样是一个类性质。

系 4.2.4

若 i 是常返的，且 $i\leftrightarrow j$，则 j 是常返的。

证明 设 m 与 n 使得 $P_{ij}^n>0$，$P_{ji}^m>0$。现在对任意的 $s\geqslant 0$ 有

$$P_{jj}^{m+n+s}\geqslant P_{ji}^m P_{ii}^s P_{ij}^n$$

从而

$$\sum_s P_{jj}^{m+n+s}\geqslant P_{ji}^m P_{ij}^n \sum_s P_{ii}^s=\infty$$

由命题 4.2.3，结论得证。

例 4.2（a） **简单随机游动**。状态空间是全体整数的集合，且转移概率

$$P_{i,i+1}=p=1-P_{i,i-1},\qquad i=0,\pm 1,\cdots$$

（其中 $0<p<1$）的马尔可夫链称为简单随机游动。这过程的一种解释是它表示一个醉汉沿一直线在游荡。另一种解释是它代表一个赌徒的输赢，在每次赌局中他或赢一元或输一元。

因为一切状态显然都相通，所以从系 4.2.4 得，它们或全是滑过的或全部是常返的。所以我们考虑状态 0，去确定 $\sum_{n=1}^{\infty}P_{00}^n$ 是有限还是无限。

因为在奇数次赌局后不可能打平（用赌博模型的解释），自然必有

$$P_{00}^{2n+1}=0,\qquad n=1,2,\cdots$$

另一方面，在 $2n$ 局之后赌徒无输赢当且仅当他赢过 n 次输过 n 次。因为在每局中以概率 p 赢得、以概率 $1-p$ 输掉，所以欲求之概率是二项概率。

$$P_{00}^{2n}=\binom{2n}{n}p^n(1-p)^n=\frac{(2n)!}{n!n!}[p(1-p)]^n,\quad n=1,2,3,\cdots$$

用斯特林近似公式：

$$n! \sim n^{n+1/2}e^{-n}\sqrt{2\pi}$$

（其中当$\lim\limits_{n\to\infty}(a_n/b_n)=1$时，我们说$a_n\sim b_n$），我们得

$$P_{00}^{2n} \sim \frac{[4p(1-p)]^n}{\sqrt{\pi n}}$$

容易验证当$a_n\sim b_n$时，$\sum\limits_n a_n<\infty$当且仅当$\sum\limits_n b_n<\infty$。因此$\sum\limits_{n=1}^{\infty}P_{00}^n$收敛当且仅当

$$\sum_{n=1}^{\infty}\frac{[4p(1-p)]^n}{\sqrt{\pi n}}$$

收敛。然而$4p(1-p)\leqslant 1$，等号成立当且仅当$p=1/2$。因此$\sum\limits_{n=1}^{\infty}P_{00}^n=\infty$当且仅当$p=\frac{1}{2}$。于是$p=\frac{1}{2}$时链是常返的，而$p\neq 1/2$时链是滑过的。

当$p=\frac{1}{2}$时，上述过程称为对称随机游动。我们也可以在多维情形中考察对称随机游动。例如，在二维对称随机游动中，每一次转移时分别以概率$\frac{1}{4}$向左、向右、向上、向下移动一步。类似地在三维时过程将以概率$\frac{1}{6}$转移到六个相邻的点中的任一个。用在一维随机游动中同样的方法可以证明二维对称随机游动是常返的，但一切更高维的随机游动是滑过的。

系 4.2.5

若$i\leftrightarrow j$，且j是常返的，则$f_{ij}=1$。

证明 假设$X_0=i$，且设n使得$P_{ij}^n>0$。若$X_n\neq j$就说我们错过了机会1。若我们错过了机会1，则以T_1记下次进入i的时刻（由系4.2.4知T_1以概率1是有限的）。若$X_{T_1+n}\neq j$就说我们错过了机会2。若错过了机会2，以T_2记下一次进入i的时刻，若$X_{T_2+n}\neq j$就说我们错过了机会3，等等。容易明白，达到首次成功所需的机会数是一个具有均值$1/P_{ij}^n$的几何随机变量，故以概率1是有限的。因为i是常返的蕴含着可能提供的机会数量是无限的，所以结论得证。

以$N_j(t)$记到时刻t为止转移到j的次数。若j是常返的，且$X_0=j$，则因为一旦转移到j，在概率上重新从头开始，故$\{N_j(t),t\geqslant 0\}$是一个来到间隔分布为$\{f_{jj}^n,n\geqslant 1\}$的更新过程。若$X_0=i$，

$i\leftrightarrow j$，且 j 是常返的，则$\{N_j(t),t\geqslant 0\}$是一个延迟更新过程，其初始来到间隔分布为$\{f_{ij}^n,\ n\geqslant 1\}$。

4.3 极限定理

容易证明，若状态 j 是滑过的，则对一切 i

$$\sum_{n=1}^{\infty}P_{ij}^n<\infty$$

这意味着从 i 出发进入状态 j 的平均次数是有限的，作为一个推论，对滑过状态 j，$n\to\infty$时，$P_{ij}^n\to 0$。

以 μ_{jj}记返回状态 j 所需的平均转移步数，即

$$\mu_{jj}=\begin{cases}\infty, & \text{若 } j \text{ 是滑过的}\\ \sum_{n=1}^{\infty}nf_{jj}^n, & \text{若 } j \text{ 是常返的}\end{cases}$$

将进入状态 j 解释为更新，从第三章的命题 3.3.1，3.3.4 及 3.4.1，我们得到下述定理。

定理 4.3.1

若 i 与 j 相通，则

(i) $P\{\lim\limits_{t\to\infty}N_j(t)/t=1/\mu_{jj}\,|\,X_0=i\}=1$；

(ii) $\lim\limits_{t\to\infty}\sum\limits_{k=1}^{n}P_{ij}^k/n=1/\mu_{jj}$；

(iii) 若 j 是非周期的，则$\lim\limits_{n\to\infty}P_{ij}^n=1/\mu_{jj}$；

(iv) 若 j 有周期 d，则$\lim\limits_{n\to\infty}P_{jj}^{nd}=d/\mu_{jj}$。

若状态 j 是常返的，则 $\mu_{jj}<\infty$时我们称它是正常返的，而 $\mu_{jj}=\infty$时称为零常返的。若我们令

$$\pi_j=\lim_{n\to\infty}P_{jj}^{nd(j)}$$

则一个常返状态 j，当 $\pi_j>0$ 时为正常返，当 $\pi_j=0$ 时为零常返，下述命题的证明留作一个练习。

命题 4.3.2

正(零)常返性是一个类性质。

一个正常返的非周期状态称为遍历的。在给出一个如何在遍历情形得到极限概率的定理之前，我们需要以下定义。

定义 对于马尔可夫链，一个概率分布 $\{P_j, j\geqslant 0\}$ 称为平稳的，若

$$P_j = \sum_{i=0}^{\infty} P_i P_{ij}, \quad j \geqslant 0$$

若 X_0 的概率分布（譬如 $P_j=P\{X_0=j\}, j\geqslant 0$）是平稳分布，则

$$P\{X_1=j\}=\sum_{i=0}^{\infty}P\{X_1=j|X_0=i\}P\{X_0=i\}$$
$$=\sum_{i=0}^{\infty}P_iP_{ij}=P_j$$

且由归纳法得

(4.3.1)
$$P\{X_n=j\}=\sum_{i=0}^{\infty}P\{X_n=j|X_{n-1}=i\}P\{X_{n-1}=i\}$$
$$=\sum_{i=0}^{\infty}P_{ij}P_i=P_j$$

因此，若初始概率分布是平稳分布，则对一切 n，X_n 有相同的分布。事实上因 $\{X_n, n\geqslant 0\}$ 是马尔可夫链，由此易得知对每个 $m\geqslant 0$，$X_n, X_{n+1}, \cdots, X_{n+m}$ 对每个 n 具有相同的分布；换言之，$\{X_n, n\geqslant 0\}$ 是一个平稳过程。

定理 4.3.3 一个不可约非周期马尔可夫链属于下列两种情况之一：

(i) 状态或全是滑过的或全是零常返的；此时对一切 i，j，$n\to\infty$ 时 $P_{ij}^n\to 0$，且不存在平稳分布。

(ii) 状态全是正常返的，即

$$\pi_j=\lim_{n\to\infty}P_{ij}^n>0$$

此时，$\{\pi_j, j=0,1,2,\cdots\}$ 是平稳分布，且不存在任何其它的平稳分布。

证明 我们先证明(ii)。首先注意，对一切 M

$$\sum_{j=0}^{M} P_{ij}^{n} \leqslant \sum_{j=0}^{\infty} P_{ij}^{n}=1$$

令 $n\to\infty$ 得，对一切 M

$$\sum_{j=0}^{M} \pi_j \leqslant 1$$

这蕴含着

$$\sum_{j=0}^{\infty} \pi_j \leqslant 1$$

现在，对一切 M

$$P_{ij}^{n+1}=\sum_{k=0}^{\infty} P_{ik}^{n} P_{kj} \geqslant \sum_{k=0}^{M} P_{ik}^{n} P_{kj}$$

令 $n\to\infty$ 得，对一切 M

$$\pi_j \geqslant \sum_{k=0}^{M} \pi_k P_{kj}$$

这蕴含着

$$\pi_j \geqslant \sum_{k=0}^{\infty} \pi_k P_{kj}, \qquad j\geqslant 0$$

为了证明上式实际上是一个等式，假设对某个 j，不等式是严格的。那么把这些不等式相加就得

$$\sum_{j=0}^{\infty} \pi_j > \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_k P_{kj}=\sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{\infty} P_{kj}=\sum_{k=0}^{\infty} \pi_k$$

这是一个矛盾。所以：

$$\pi_j=\sum_{k=0}^{\infty} \pi_k P_{kj}, \qquad j=0,1,2,\cdots$$

记 $P_j=\pi_j \Big/ \sum_{k=0}^{\infty} \pi_k$，可见 $\{P_j, j=0,1,2,\cdots,\}$ 是一个平稳分布，因此至少有一平稳分布存在，今设 $\{P_j, j=0,1,2\cdots\}$ 是任一平稳分布。若 $\{P_j, j=0,1,2,\cdots\}$ 是 X_0 的概率分布，则由(4.3.1)有

$$\begin{aligned} P_j &= P\{X_n=j\} \\ &= \sum_{i=0}^{\infty} P\{X_n=j \mid X_0=i\} P\{X_0=i\} \qquad (4.3.2) \\ &= \sum_{i=0}^{\infty} P_{ij}^{n} P_i \end{aligned}$$

从(4.3.2)看出，对一切 M

$$P_j \geqslant \sum_{i=0}^{M} P_{ij}^n P_i$$

先令 n 后令 M 趋于∞得

(4.3.3) $$P_j \geqslant \sum_{i=0}^{\infty} \pi_j P_i = \pi_j$$

为走另一条路证明 $P_j \leqslant \pi_j$，用(4.3.2)及 $P_{ij}^n \leqslant 1$ 得出，对一切 M

$$P_j \leqslant \sum_{i=0}^{M} P_{ij}^n P_i + \sum_{i=M+1}^{\infty} P_i$$

令 $n \to \infty$得，对一切 M

$$P_j \leqslant \sum_{i=0}^{M} \pi_j P_i + \sum_{i=M+1}^{\infty} P_i$$

因$\sum_{i=0}^{\infty} P_i = 1$，所以，令 $M \to \infty$即得

(4.3.4) $$P_j \leqslant \sum_{i=0}^{\infty} \pi_j P_i = \pi_j$$

若诸状态是滑过的或零常返的，且$\{P_j, j=0,1,2,\cdots\}$是一平稳分布，则等式(4.3.2)成立，且 $P_{ij}^n \to 0$，这显然是不可能的。于是对于情形 (i)，不存在任何平稳分布，证毕。

注记

(1) 当情况如定理 4.3.3 (ii)所述时，我们称马尔可夫链是遍历的。

(2)若过程以极限概率开始,则所得马尔可夫链是平稳的,这是十分直观的。因为此时马尔可夫链处在时刻 0 等价于一个具有相同 P 矩阵的独立马尔可夫链处在时刻∞。因此原来的马尔可夫链处在时刻 t 等价于第二个链处在时刻$\infty + t = \infty$,所以是平稳的。

(3) 在不可约，正常返，周期情形，仍有 π_j，$j \geqslant 0$，是

$$\pi_j = \sum_{i=0}^{\infty} \pi_j P_{ij}$$

$$\sum_{j=0}^{\infty} \pi_j = 1$$

的唯一非负解。但现在 π_j 必须被解释为马尔可夫链长时间之后处于状态 j 的时间所占的比率(见习题 14)。于是，$\pi_j = 1/\mu_{jj}$，而由

定理 4.3.1(iv)可知，跨 $nd(j)$ 的步子从 j 到 j 的极限概率为

$$\lim_{n\to\infty}P_{jj}^{nd}=\frac{d}{\mu_{jj}}=d\pi_j$$

其中 d 是马尔可夫链的周期。

例 4.3(a)　$M/G/1$ 的嵌入链的极限概率。 考虑例 4.1(a)中 $M/G/1$ 系统的嵌入马尔可夫链，设

$$a_j=\int_0^\infty e^{-\lambda x}\frac{(\lambda x)^j}{j!}dG(x)$$

即 a_j 是在一个服务期内有 j 个顾客来到的概率，这个链的转移概率是

$$\begin{aligned}&P_{0j}=a_j,\\&P_{ij}=a_{j-i+1},\quad i>0,\quad j\geqslant i-1\\&P_{ij}=0,\qquad j<i-1\end{aligned}$$

设 $\rho=\sum_{j=0}^{\infty}ja_j$。由于 ρ 等于在一个服务期内来到的顾客的平均数，对服务期的长度取条件可得

$$\rho=\lambda E[S]$$

其中 S 是服务时期，具有分布 G。

现在我们证明当 $\rho<1$ 时马尔可夫链是正常返的，为此解方程组

$$\pi_j=\sum_{i=0}^{\infty}\pi_iP_{ij}$$

现在这些方程的形式为

(4.3.5)
$$\pi_j=\pi_0a_j+\sum_{i=1}^{j+1}\pi_ia_{j-i+1},\qquad j\geqslant 0$$

为求解，我们引进母函数

$$\pi(s)=\sum_{j=0}^{\infty}\pi_js^j,\qquad A(s)=\sum_{j=0}^{\infty}a_js^j$$

以 s^j 乘(4.3.5)的两边，且对 j 求和得

$$\begin{aligned}\pi(s)&=\pi_0A(s)+\sum_{j=0}^{\infty}\sum_{i=1}^{j+1}\pi_ia_{j-i+1}s^j\\&=\pi_0A(s)+s^{-1}\sum_{i=1}^{\infty}\pi_is^i\sum_{j=i-1}^{\infty}a_{j-i+1}s^{j-i+1}\\&=\pi_0A(s)+(\pi(s)-\pi_0)A(s)/s\end{aligned}$$

或

$$\pi(s)=\frac{(s-1)\pi_0 A(s)}{s-A(s)}$$

为了计算 π_0，在上式中令 $s\to 1$。因为

$$\lim_{s\to 1} A(s)=\sum_{i=0}^{\infty} a_i=1$$

由此得

$$\lim_{s\to 1}\pi(s)=\pi_0\lim_{s\to 1}\frac{s-1}{s-A(s)}$$
$$=\pi_0(1-A'(1))^{-1}$$

其中最后的等式由洛必达法则而得。今有

$$A'(1)=\sum_{i=0}^{\infty} i a_i=\rho$$

于是

$$\lim_{s\to 1}\pi(s)=\frac{\pi_0}{1-\rho}$$

然而，因为 $\lim_{s\to 1}\pi(s)=\sum_{i=0}^{\infty}\pi_i$，推知 $\sum_{i=0}^{\infty}\pi_i=\pi_0/(1-\rho)$；于是平稳概率存在当且仅当 $\rho<1$，且此时有

$$\pi_0=1-\rho=1-\lambda E[S]$$

因此，当 $\rho<1$，或等价地，$E[S]<1/\lambda$ 时，有

$$\pi(s)=\frac{(1-\lambda E[s])(s-1)A(s)}{s-A(s)}$$

例 4.3(b)　$G/M/1$ 的嵌入链的极限概率。考虑例 4.1(b) 中所述 $G/M/1$ 排队系统的嵌入马尔可夫链。极限概率 π_k，$k=0, 1, \cdots$，可作为下列方程的唯一解而得到

$$\pi_k=\sum_{i=0}^{\infty}\pi_i P_{ik},\qquad k\geqslant 0$$

$$\sum_{k=0}^{\infty}\pi_k=1$$

此时它们归结为

$$(4.3.6)\qquad \pi_k=\sum_{i=k-1}^{\infty}\pi_i\int_0^{\infty}e^{-\mu t}\frac{(\mu t)^{i+1-k}}{(i+1-k)!}dG(t),\qquad k\geqslant 1$$

$$\sum_{k=0}^{\infty}\pi_k=1$$

(我们未包括方程 $\pi_0=\sum_{i=0}^{\infty}\pi_i P_{i0}$，因为总有一个方程是多余的。)

为了解上列方程，我们试求形成为 $\pi_k=c\beta^k$ 的解。代入(4.3.6)得

(4.3.7) $$c\beta^k=c\sum_{i=k-1}^{\infty}\beta^i\int_0^{\infty}e^{-\mu t}\frac{(\mu t)^{i+1-k}}{(i+1-k)!}dG(t)$$
$$=c\int_0^{\infty}e^{-\mu t}\beta^{k-1}\sum_{i=k-1}^{\infty}\frac{(\beta\mu t)^{i+1-k}}{(i+1-k)!}dG(t)$$
$$=c\int_0^{\infty}e^{-\mu t}\beta^{k-1}e^{\beta\mu t}dG(t)$$

或

(4.3.8) $$\beta=\int_0^{\infty}e^{-\mu t(1-\beta)}dG(t)$$

常数 c 从 $\sum_{k=0}^{\infty}\pi_k=1$ 可得：

$$c=1-\beta$$

因为 π_k 是(4.3.6)的唯一解且 $\pi_k=(1-\beta)\beta^k$ 满足它，故得

$$\pi_k=(1-\beta)\beta^k,\qquad k=0,1,\cdots$$

其中 β 是方程(4.3.8)的解。(能证明若 G 的均值大于平均服务时间 $1/\mu$，则存在唯一的 β 满足(4.3.8)，它介于0与1之间。)β 的精确值通常只能用数值方法得到。

例4.3(c)　更新过程的年龄。 开始时一个部件投入使用，当它失效时，在下一个使用期的初始时刻就被另一个新的部件替换。假设各部件的寿命是独立的，且在第 i 个使用期中失效的概率为 $P_i, i\geqslant 1$，其中分布 $\{P_i\}$ 是非周期的，且 $\sum_{i=0}^{\infty}iP_i<\infty$。以 X_n 记在时刻 n 正在使用中的部件的年龄即使用期数(包括第 n 个使用期)。若以

$$\lambda(i)=\frac{P_i}{\sum_{j=i}^{\infty}P_j}$$

记年龄为 i 的部件失效的概率，则 $\{X_n, n\geqslant 0\}$ 是马尔可夫链，其转移概率为

$$P_{i,1}=\lambda(i)=1-P_{i,i+1},\qquad i\geqslant 1$$

因此极限概率满足

(4.3.9) $$\pi_1=\sum_{i=0}^{\infty}\pi_i\lambda(i)$$

(4.3.10) $$\pi_{i+1}=\pi_i(1-\lambda(i)),\qquad i\geqslant 1$$

叠代(4.3.10)得

$$\begin{aligned}\pi_{i+1}&=\pi_i(1-\lambda(i))\\&=\pi_{i-1}(1-\lambda(i))(1-\lambda(i-1))\\&=\pi_1(1-\lambda(1))(1-\lambda(2))\cdots(1-\lambda(i))\\&=\pi_1\sum_{j=i+1}^{\infty}P_j\\&=\pi_1P\{X\geqslant i+1\}\end{aligned}$$

其中 X 是一个部件的寿命。利用 $\sum_{i=0}^{\infty}\pi_i=1$ 得

$$1=\pi_1\sum_{i=0}^{\infty}P\{X\geqslant i\}$$

或

$$\pi_1=1/E[X]$$

且

(4.3.11) $$\pi_i=P\{X\geqslant i\}/E[X],\qquad i\geqslant 1$$

易见它满足 (4.3.9)。

值得注意的是,(4.3.11)是在意料之中的,因为在非格点情形年龄的极限分布是平衡分布(见第三章 3.4 节),其密度是 $\overline{F}(x)/E[X]$。

4.4 类之间的转移与赌徒输光问题

在本节中我们先证明常返类是闭的,即一旦进入就永不离开。

命题 4.4.1

设 R 是一个常返的状态类。若 $i\in R$,$j\notin R$,则 $P_{ij}=0$。

证明 假设 $P_{ij}>0$。那么,因为 i 与 j 不相通(因 $j\notin R$),所以对一切 n,$P_{ji}^n=0$。因此若过程从状态 i 出发,则存在一至少为 P_{ij} 的正概率使过程永不回到 i,此与 i 是常返的相矛盾,从而有 $P_{ij}=0$。

设 j 是一给定的常返状态,以 T 记全部滑过状态的集合,对 $i\in T$,我们经常感兴趣的是计算 f_{ij},过程从 i 出发迟早到达 j 的概率。下述命题对初次转移后的状态取条件得出一组 f_{ij} 满足的方程。

命题 4.4.2 若 j 是常返的，则概率组 $\{f_{ij}, i\in T\}$ 满足

$$f_{ij}=\sum_{k\in T}P_{ik}f_{kj}+\sum_{k\in R}P_{ik}, \qquad i\in T$$

其中 R 记与 j 相通的状态的集合。

证明

$$\begin{aligned}f_{ij}&=P\{N_j(\infty)>0|X_0=i\}\\&=\sum_{一切k}P\{N_j(\infty)>0|X_0=i,X_1=k\}P\{X_1=k|X_0=i\}\\&=\sum_{k\in T}f_{kj}P_{ik}+\sum_{k\in R}f_{kj}P_{ik}+\sum_{\substack{k\notin R\\k\notin T}}f_{kj}P_{ik}\\&=\sum_{k\in T}f_{kj}P_{ik}+\sum_{k\in R}P_{ik}\end{aligned}$$

其中我们用系 4.2.5 断言 $k\in R$ 时 $f_{kj}=1$ 及用命题 4.4.1 断言 $k\notin T, k\notin R$ 时，$f_{kj}=0$。

例 4.4 (a)　赌徒输光问题。考虑一赌徒，在每局赌博中他以概率 p 赢一元以概率 $q=1-p$ 输一元，假定各局赌博是独立的，赌徒开始有 i 元，问他的赌金在到达 0(输光)之前达到 N 元的概率是多少。

以 X_n 记赌徒在时刻 n 的赌金，则过程 $\{X_n, n=0,1,2,\cdots,\}$ 是马尔可夫链，其转移概率为

$$P_{00}=P_{NN}=1$$

$$P_{i,i+1}=p=1-P_{i,i-1}, \quad i=1,2,\cdots,N-1$$

此马尔可夫链有三个类，即 $\{0\}$，$\{1,2,\cdots,N-1\}$ 与 $\{N\}$，第一与第三个是常返的，而第二个是滑过的，由于滑过状态只能到达有限次，所以在有限时间后赌徒或将达到 N 元的目标或输光。

以 $f_i\equiv f_{iN}$ 记赌徒从 i 元的赌本开始，$0\leqslant i\leqslant N$，最终达到 N 的概率。对第一局赌博的结果取条件(或等价地，用命题 4.4.2)，得

$$f_i=pf_{i+1}+qf_{i-1}, \qquad i=1,2,\cdots N-1$$

或等价地，因 $p+q=1$

$$f_{i+1}-f_i=\frac{q}{p}(f_i-f_{i-1}), \qquad i=1,2,\cdots,N-1$$

由于 $f_0=0$，从上式可见

$$f_2-f_1=\frac{q}{p}(f_1-f_0)=(q/p)f_1$$

$$f_3-f_2=\frac{q}{p}(f_2-f_1)=(q/p)^2f_1$$

$$\vdots$$

$$f_i - f_{i-1} = \frac{q}{p}(f_{i-1} - f_{i-2}) = (q/p)^{i-1} f_1$$

$$\vdots$$

$$f_N - f_{N-1} = \frac{q}{p}(f_{N-1} - f_{N-2}) = (q/p)^{N-1} f_1$$

将前 $i-1$ 个方程相加得

$$f_i - f_1 = f_1[(q/p) + (q/p)^2 + \cdots + (q/p)^{i-1}]$$

或

$$f_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)} f_1, & \text{若 } q/p \neq 1 \\ i f_1, & \text{若 } q/p = 1 \end{cases}$$

利用 $f_N = 1$ 得

$$f_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)^N}, & \text{若 } p \neq 1/2 \\ i/N, & \text{若 } p = 1/2 \end{cases}$$

有趣的是注意到，$N \to \infty$ 时有

$$f_i \to \begin{cases} 1-(q/p)^i, & \text{若 } p > 1/2 \\ 0, & \text{若 } p \leqslant 1/2 \end{cases}$$

因此由概率的连续性可得，在与有无穷赌本的对手赌博中，当 $p>1/2$ 时，赌徒的赌本以一正概率趋于无穷，而当 $p \leqslant 1/2$ 时，将以概率 1 输光。

4.5 分支过程

考虑一个由能产生同类后代的个体组成的群体。每一个体生命结束时以概率 P_j，$j \geqslant 0$，产生了 j 个新的后代，与别的个体产生的后代个数相互独立。初始的个体数以 X_0 表示，称为第零代的总数，第零代的后代构成第一代，其总数记作 X_1。一般地，以 X_n 记第 n 代的总数。此马尔可夫链 $\{X_n, n \geqslant 0\}$ 称为分支过程。

假设 $X_0 = 1$，注意到

$$X_n = \sum_{i=1}^{X_{n-1}} Z_i$$

其中 Z_i 表示第 $n-1$ 代的第 i 个个体的后代个数，我们可以计算 X_n 的均值。对 X_{n-1} 取条件得

$$\begin{aligned}E[X_n]&=E[E[X_n|X_{n-1}]]\\&=\mu E[X_{n-1}]\\&=\mu^2E[X_{n-2}]\\&=\mu^n\end{aligned}$$

其中 μ 是每个个体的后代个数的均值。

以 π_0 记从单个个体开始群体迟早灭绝的概率。

对初始个体的后代个数取条件，可以导出一个确定 π_0 的方程如下：

$$\begin{aligned}\pi_0&=P\{\text{群体灭绝}\}\\&=\sum_{j=0}^{\infty}P\{\text{群体灭绝}|X_1=j\}P_j\end{aligned}$$

给定 $X_1=j$，群体最终灭绝当且仅当以第一代的成员为始祖的 j 个家族最终灭绝。由于各家族假定是独立活动的，而任一家族灭绝的概率都是 π_0，所以得到

(4.5.1) $$\pi_0=\sum_{j=0}^{\infty}\pi_0^jP_j$$

事实上可证明下列

定理 4.5.1 设 $P_0>0$ 及 $P_0+P_1<1$，则 (i) π_0 是满足

$$\pi_0=\sum_{j=0}^{\infty}\pi_0^jP_j$$

的最小正数。(ii) $\pi_0=1$ 的充要条件是 $\mu\leqslant1$。

证明 为了证明 π_0 是(4.5.1)的最小解，设 $\pi\geqslant0$ 满足(4.5.1)。首先我们用归纳法证明，对于一切 n，$\pi\geqslant P\{X_n=0\}$。现有

$$\pi=\sum_{j=0}^{\infty}\pi^jP_j\geqslant\pi^0P_0=P_0=P\{X_1=0\}$$

假定 $\pi\geqslant P\{X_n=0\}$。则

$$P\{X_{n+1}=0\}=\sum_{j=0}^{\infty}P\{X_{n+1}=0|X_1=j\})P_j$$

$$
\begin{aligned}
&= \sum_{j=0}^{\infty} (P\{X_n = 0\})^j P_j \\
&\leqslant \sum_{j=0}^{\infty} \pi^j P_j \qquad \text{(由归纳法假设)} \\
&= \pi
\end{aligned}
$$

因此，对一切 n

$$\pi \geqslant P\{X_n = 0\}$$

令 $n \to \infty$

$$\pi \geqslant \lim_{n\to\infty} P\{X_n = 0\} = P\{\text{群体灭绝}\} = \pi_0$$

为了证明(ii)，定义母函数

$$\phi(s) = \sum_{j=0}^{\infty} s^j P_j$$

因为 $P_0 + P_1 < 1$，对一切 $s \in (0, 1)$

$$\phi''(s) = \sum_{j=0}^{\infty} j(j-1)s^{j-2}P_j > 0$$

因此 $\phi(s)$ 在开区间 $(0,1)$ 中是严格凸的函数，现在分两种情形(图 4.5.1 与图 4.5.2)。在图 4.5.1 中，对一切 $s \in (0,1)$，$\phi(s) > s$，而在图 4.5.2 中，对某个 $s \in (0,1)$，$\phi(s) = s$。几何上很明显，图 4.5.1 表示 $\phi'(1) \leqslant 1$ 时的图形，而图 4.5.2 适合于 $\phi'(1) > 1$ 的情形。于是，因 $\phi(\pi_0) = \pi_0$，$\pi_0 = 1$ 当且仅当 $\phi'(1) \leqslant 1$。由于 $\phi'(1) = \sum_{j=1}^{\infty} jP_j = \mu$，结果得证。

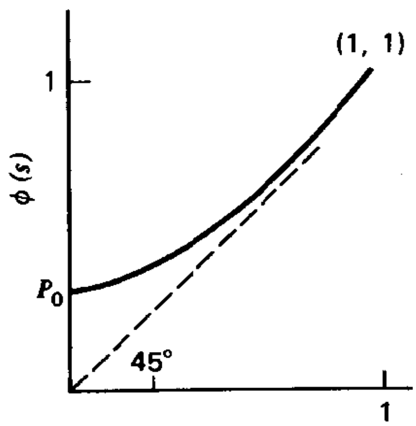


图 4.5.1

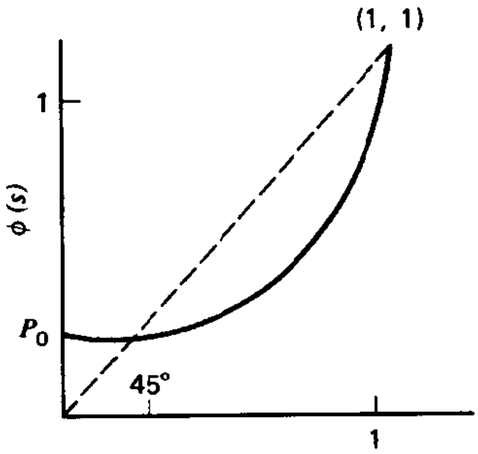


图 4.5.2

4.6 马尔可夫链的应用

4.6.1 算法有效性的马尔可夫链模型

在运筹学与计算机科学中，某些算法以如下的方式进行：目标是要决定 N 个有序元素中的最优者，算法以其中一个元素开始，而后逐次移动到更好的元素，直到到达最优者为止。（最重要的例子可能是线性规划的单纯形算法，该算法试图要求一个受线性约束条件的线性函数的最大值，此时一个元素对应于可行区域的一个端点。）如果从“最差的情况”的角度考察算法的效率，那么一般都能构造出大致需 $N-1$ 步才到达最优元素的例子。本节中我们将对必需的步数给出一个简单的概率模型。特别地，我们考虑一个马尔可夫链，它从任何状态等可能地进入到任一更好的状态。

考虑一个马尔可夫链，其 $P_{11}=1$，而

$$P_{ij}=\frac{1}{i-1},\qquad j=1,\cdots,i-1,\quad i>1$$

以 T_i 记以状态 i 转移到状态 1 的步数，对初始转移取条件可得 $E[T_i]$ 的一个递推公式：

$$E[T_i]=1+\frac{1}{i-1}\sum_{j=1}^{i-1}E[T_j] \tag{4.6.1}$$

从 $E[T_1]=0$ 开始，逐一看到

$$E[T_2]=1$$

$$E[T_3]=1+1/2$$

$$E[T_4]=1+\frac{1}{3}\left(1+1+\frac{1}{2}\right)=1+\frac{1}{2}+\frac{1}{3}$$

不难猜想，且可用归纳法证明

$$E[T_i]=\sum_{j=1}^{i-1}\frac{1}{j}$$

然而，为了更完全地描述 T_N，我们利用表达式

$$T_N = \sum_{j=1}^{N-1} I_j$$

其中

$$I_j = \begin{cases} 1, & \text{若过程曾进入 } j \\ 0, & \text{其它} \end{cases}$$

上述表达式的重要性来自下列

引理 4.6.1

$I_1, \cdots, I_N$ 是独立的，且

$$P\{I_j = 1\} = 1/j, \qquad 1 \leqslant j \leqslant N-1$$

证明 给定 $I_{j+1}, \cdots, I_N$，令 $n = \min\{i: i>j, I_i=1\}$，则

$$P\{I_j = 1 | I_{j+1}, \cdots, I_N\} = \frac{1/(n-1)}{j/(n-1)} = \frac{1}{j}$$

命题 4.6.2

(i) $E[T_N] = \sum_{j=1}^{N-1} \frac{1}{j}$

(ii) $\mathrm{Var}(T_N) = \sum_{j=1}^{N-1} \frac{1}{j}\left(1 - \frac{1}{j}\right)$

(iii) 对充分大的 N，T_N 渐近泊松分布，均值为 $\log N$。

证明 从引理 4.6.1 与表达式 $T_N = \sum_{j=1}^{N-1} I_j$ 可得(i)及(ii)。大量独立的贝努里随机变量的和，当各项不为零的概率很小时，是渐近泊松分布的。因为

$$\int_1^N \frac{dx}{x} < \sum_{j=1}^{N-1} \frac{1}{j} < 1 + \int_1^{N-1} \frac{dx}{x}$$

或

$$\log N < \sum_{j=1}^{N-1} \frac{1}{j} < 1 + \log(N-1)$$

从而

$$\log N \approx \sum_{j=1}^{N-1} \frac{1}{j}$$

所以得(iii)。

4.6.2 应用于游程——一个连续状态空间的马尔可夫链

考虑一列数 x_1，x_2，…。如果我们在 x_1 之前立一竖线，每当 $x_j>x_{j+1}$ 就在 x_j 与 x_{j+1} 之间立一竖线，那么两条竖线之间的段落称为游程。例如，部分序列 3，5，8，2，4，3，1 包括四个游程，如下所示

$$|\ 3,\ 5,\ 8\ |\ 2,\ 4\ |\ 3\ |1$$

这样，每个游程是序列的一个上升的段落。

现假设 $X_1, X_2, \cdots$ 是独立同分布的(0,1)上均匀分布的随机变量，我们感兴趣的是逐个游程的长度的分布，例如以 L_1 记初始的游程的长度，那么 L_1 至少是 m 当且仅当前 $m-1$ 个值是依次上升的，可见

$$P\{L_1 \geqslant m\} = \frac{1}{(m-1)!}, \quad m = 1,2,\cdots$$

其后的游程的分布也容易得到，如果我们知道这一游程的初始值 x。因为若一游程起始于 x，则

$$P\{L \geqslant m | x\} = \frac{(1-x)^{m-1}}{(m-1)!} \tag{4.6.2}$$

因为要游程长至少是 m，其后的 $m-1$ 个值都必须大于 x 且必须依次增大。

为了得到一个给定的游程长度的无条件分布，以 I_n 记第 n 个游程的初始值。容易看到 $\{I_n, n\geqslant 1\}$ 是一具有连续状态空间的马尔可夫链，为了计算 $p(y|x)$，在一个游程从初始值 x 开始的条件下，下一游程起始于 y 的概率密度，作如下的推导：

$$\begin{aligned} &P\{I_{n+1} \in (y, y+dy) | I_n = x\} \\ &\quad = \sum_{m=1}^{\infty} P\{I_{n+1} \in (y, y+dy), L_n = m | I_n = x\} \end{aligned}$$

其中 L_n 是第 n 个游程之长。此游程长为 m，且下一游程起始于 y，如果

(i) 其后的 $m-1$ 个值是依次增加且全部大于 x；

(ii) 第 m 个值必须等于 y；

(iii) 前 $m-1$ 个值的最大者必超过 y。

因此，

$$P\{I_{n+1}\in(y,y+dy),L_n=m|I_n=x\}$$
$$=\frac{(1-x)^{m-1}}{(m-1)!}dyP\{\max(X_1,\ \cdots X_{m-1})>y|X_i>x,\quad i=1,\ \cdots,\ m-1\}$$
$$=\begin{cases}\dfrac{(1-x)^{m-1}}{(m-1)!}dy, & 若\ y<x\\ \dfrac{(1-x)^{m-1}}{(m-1)!}dy\left[1-\left(\dfrac{y-x}{1-x}\right)^{m-1}\right], & 若\ y>x\end{cases}$$

对 m 求和得

$$p(y|x)=\begin{cases}e^{1-x}, & 若\ y<x\\ e^{1-x}-e^{y-x}, & 若\ y>x\end{cases}$$

所以，$\{I_n,\ n\geqslant 1\}$是一马尔可夫链，具有连续状态空间(0,1)且转移概率密度由上式给定。

为了得到 I_n 的极限分布，我们先冒险作一个猜测，而后利用类似于定理 4.3.3 的结果来验证我们的猜测。I_1 是第一个游程的初始值，在(0,1)上均匀分布。然而，只要有一个值小于其前一值，后一游程便开始。因而看来有理由认为，小于 y 的这样的值所占的比例长时间之后要等于一个(0,1)上均匀分布的随机变量在小于第二个相互独立的(0,1)上均匀分布变量的条件下，小于 y 的概率。由于

$$P\{X_2>y|X_2<X_1\}=\frac{\frac{1}{2}(1-y)^2}{1/2}=(1-y)^2$$

所以，I_n 的极限密度 $\pi(y)$等于

$$\pi(y)=2(1-y),\qquad 0<y<1$$

似乎是合理的。[此极限密度的第二种直接的推导如下：每一个取值 y 的 X_i 将是一个新游程的起点，如果它的前一个值大于 y。因此初始值在$(y,y+dy)$中的游程发生的速率看来等于 $\overline{F}(y)f(y)dy=(1-y)dy$，所以初始值在$(y,y+dy)$中的游程所占比例

将是 $(1-y)dy/\int_0^1(1-y)dy=2(1-y)dy$。]因为可对连续状态空间的马尔可夫链证明一个类似于定理 4.3.3 的定理，所以为了证明上面的猜测，我们需验证：

$$\pi(y)=\int_0^1\pi(x)p(y|x)dx$$

这时就归结为证明

$$1-y=\int_0^y(e^{1-x}-e^{y-x})(1-x)dx+\int_y^1 e^{1-x}(1-x)dx$$

或

$$1-y=\int_0^1 e^{1-x}(1-x)dx-\int_0^y e^{y-x}(1-x)dx$$

只要利用下列恒等式很容易证得此式：

$$\int ze^z dz=ze^z-e^z$$

这样，我们已证明第 n 个游程的初始值 I_n 的极限密度是 $\pi(x)=2(1-x)$，$0<x<1$。因此第 n 个游程的长度 L_n 的极限分布为

$$\lim_{n\to\infty}P\{L_n\geqslant m\}=\int_0^1\frac{(1-x)^{m-1}}{(m-1)!}2(1-x)dx \tag{4.6.3}$$

$$=\frac{2}{(m+1)(m-1)!}$$

为了计算一个游程的平均长度，注意到由(4.6.2)有

$$E[L|I=x]=\sum_{m=1}^{\infty}\frac{(1-x)^{m-1}}{(m-1)!}=e^{1-x}$$

从而

$$\lim_{n\to\infty}E[L_n]=\int_0^1 e^{1-x}2(1-x)dx$$

$$=2$$

上述极限也可从(4.6.3)算得如下：

$$\lim_{n\to\infty}E[L_n]=2\sum_{m=1}^{\infty}\frac{1}{(m+1)(m-1)!}$$

它导致有趣的恒等式

$$1=\sum_{m=1}^{\infty}\frac{1}{(m+1)(m-1)!}$$

4.6.3 名册排序规则——移前一位规则的最优性

假设给定 n 个元素 $e_1, \cdots, e_n$，将它们按某种顺序排列，每次要求取出其中一个元素，然后放回并将 n 个元素重新排序，每次取出 e_i(与过去独立)的概率为 $P_i, P_i \geqslant 0, \sum_{i=1}^{n} P_i = 1$。感兴趣的问题是要确定一种最优排序规则以便使长时间后被取出的元素的平均位置最小。显然若 P_i 已知，则最优的排序规则就是每次都按照 P_i 的递减次序去安排诸元素。事实上，即使诸 P_i 未知，我们也能渐近地做到，即每次依照各元素已被取出的次数的递减次序去排列它们。然而，问题会变得更有趣，如果我们不允许有上述排序规则所必要的存储记忆的功能，而对重新排序的规则作如下的限制：每次元素的重新排序只允许依赖于现在的排列次序及这次取出的元素的位置。

对于一个给定的重新排序规则，被取出的元素的平均位置至少在理论上可以通过分析有 $n!$ 个状态的马尔可夫链得到，这个马尔可夫链在任一时刻的状态就是元素这时的排列次序。然而，对这么多的状态，分析很快变成不可行的，所以我们把问题简化，假定概率满足：

$$P_1 = p, \quad P_2 = \cdots = P_n = \frac{1-p}{n-1} = q$$

这时，因为从第二列第 n 个元素，在被取出的概率相等的意义下是无区别的，被取出的元素的平均位置可以通过分析简单得多的、只有 n 个状态的马尔可夫链得到，这马尔可夫链的状态就是 P_1 的位置。在对概率作这样的假设之下，在一大类规则之中，把取出的元素移到前面一个位置上去的“移前一位规则”是最优的。

考虑如下限定的一类规则，当取出的元素是在位置 i 上时，把此元素移到位置 j_i 上去，而其它的元素的相对位置保持不动。此

外假设，对 $i>1$，$j_1=1$ 及 $j_i\geqslant j_{i-1}$，$i=2$，3，…，n 有 $j_i<i$。集合$\{j_i,i=1,2,\cdots,n\}$就刻划了一个规则。

对于给定的一个此类规则，令

$$K(i)=\max\{l:j_{i+l}\leqslant i\}$$

换句话说，对任意的 i，任何位置 i，$i+1$，…，$i+K(i)$上的一个元素如果被取出，将被移到小于或等于 i 的位置上。

对于一个特定的上述类型的规则 R(譬如说对于它 $K(i)=k(i)$， $i=1,2,\cdots,n$)，在使用此规则时的平稳分布记为：

$$\pi_i=P_\infty\{e_1\text{ 处于位置 }i\},\quad i\geqslant 1$$

此外，令

$$\Pi_i=\sum_{j=i+1}^{n}\pi_j=P_\infty\{e_1\text{ 的位置大于 }i\},\quad i\geqslant 0$$

记号 P_∞意味着上述概率是极限概率。在写下稳态方程之前，值得注意以下事实：

(i) 任何元素每次至多朝后移动一个位置。

(ii) 如果一个元素处于位置 i，而且既不是它也不是它后面的 $k(i)$个位置上的元素被取出，则它将仍然在位置 i 上。

(iii) 任何处于诸位置 $i,i+1,\cdots,i+k(i)$上的元素若被取出，它将被移到$\leqslant i$ 的位置上。

现在容易看到稳态概率是：

$$\Pi_i=\Pi_{i+k(i)}+(\Pi_i-\Pi_{i+k(i)})(1-p)+(\Pi_{i-1}-\Pi_i)qk(i)$$

此式成立是因为元素 e_1 将在大于 i 的位置上，如果它在前一次的位置大于 $i+k(i)$，或者它的位置小于 $i+k(i)$且大于 i 然而未被取出，或者它在位置 i 上且取出了位置 $i+1$，…，$i+k(i)$上的一个元素。上面的方程等价于

$$\text{(4.6.4)}\qquad \begin{aligned}&\Pi_i=a_i\Pi_{i-1}+(1-a_i)\Pi_{i+k(i)},\quad i=1,\cdots,n-1\\&\Pi_0=1,\quad \Pi_n=0\end{aligned}$$

其中

$$a_i=\frac{qk(i)}{qk(i)+p}$$

现在考虑一个特殊的属于所限制的类型的规则，即移前一位规则，它有 $j_i=i-1$，$i=2$，…，n，$j_1=1$，对移前一位规则，相应的 Π_i 记为 $\overline{\Pi}_i$。对此规则 $K(i)=1$，所以从方程(4.6.4)得

$$\overline{\Pi}_i=\frac{q\overline{\Pi}_{i-1}+p\overline{\Pi}_{i+1}}{p+q}$$

或等价地

$$\overline{\Pi}_{i+1}-\overline{\Pi}_i=\frac{q}{p}(\overline{\Pi}_i-\overline{\Pi}_{i-1})$$

这蕴含着

$$\begin{aligned}\overline{\Pi}_{i+j}-\overline{\Pi}_{i+j-1}&=\frac{q}{p}(\overline{\Pi}_{i+j-1}-\overline{\Pi}_{i-j-2})\\&\;\vdots\\&=\left(\frac{q}{p}\right)^{j}(\overline{\Pi}_i-\overline{\Pi}_{i-1})\end{aligned}$$

把上面的方程对 $j=1$，…，r 相加得

$$\overline{\Pi}_{i+r}-\overline{\Pi}_i=(\overline{\Pi}_i-\overline{\Pi}_{i-1})\left[\frac{q}{p}+\cdots+\left(\frac{q}{p}\right)^{r}\right],\quad i+r\leqslant n$$

令 $r=k(i)$,其中 $k(i)$是我们所讨论的规则类中的一个给定的规则 R 的 $K(i)$的值,我们看到

$$\overline{\Pi}_{i+k(i)}-\overline{\Pi}_i=(\overline{\Pi}_i-\overline{\Pi}_{i-1})\left[\frac{q}{p}+\cdots+\left(\frac{q}{p}\right)^{k(i)}\right]$$

或等价地

(4.6.5) $$\begin{aligned}&\overline{\Pi}_i=b_i\overline{\Pi}_{i-1}+(1-b_i)\overline{\Pi}_{i+k(i)},\quad i=1,\cdots,n-1\\&\overline{\Pi}_0=1,\qquad \overline{\Pi}_n=0\end{aligned}$$

其中

$$b_i=\frac{(q/p)+\cdots+(q/p)^{k(i)}}{1+(q/p)+\cdots+(q/p)^{k(i)}},\qquad i=1,\cdots,n-1$$

现在我们可以证明以下的

命题 4.6.3

若 $p\geqslant 1/n$，则对一切 i $\overline{\Pi}_i\leqslant\Pi_i$。

若 $p\leqslant 1/n$，则对一切 i $\overline{\Pi}_i\geqslant\Pi_i$。

证明 考虑 $p\geqslant 1/n$ 的情形，它等价于 $p\geqslant q$，注意此时有

$$a_i=1-\frac{1}{1+(k(i)/p)q}\geqslant 1-\frac{1}{1+q/p+\cdots+(q/p)^{k(i)}}=b_i$$

现在定义一马尔可夫链具有状态 0，1，…，n 及转移概率

$$P_{0,0}=P_{n,n}=1$$

$$P_{ij}=\begin{cases}c_i & 若\ j=i-1,\\ 1-c_i & 若\ j=i+k(i),\end{cases}\quad i=1,\cdots,n-1 \tag{4.6.6}$$

以 f_i 记马尔可夫链以状态 i 出发迟早进入状态 0 的概率，则 f_i 满足

$$f_i=c_if_{i-1}+(1-c_i)f_{i+k(i)},\quad i=1,\cdots,n-1$$

$$f_0=1,\quad f_n=0$$

能够证明上列方程组有唯一解，所以，如果我们对一切 i 取 c_i 等于 a_i 则由（4.6.4）f_i 将等于规则 R 的 Π_i；如果我们令 $c_i=b_i$，则由（4.6.5）f_i 等于 $\overline{\Pi}_i$，由（4.6.6）定义的马尔可夫链迟早进入状态 0 的概率是向量 $\underline{c}=(c_1,\cdots,c_{n-1})$ 的一个递增函数，这在直观上是清楚的（我们推迟到第八章再给出一个正式的证明）。因为 $a_i\geqslant b_i$，$i=1,\cdots,n$，我们看到对于一切 i

$$\Pi_i\geqslant\overline{\Pi}_i$$

当 $p\leqslant 1/n$ 时，$a_i\leqslant b_i$，$i=1,\cdots,n-1$，从而上面不等式反向。

定理 4.6.4 在所考虑的规则中，移前一位规则使被取出的元素的极限平均位置最小。

证明 以 X 记 e_1 的位置，以 e_1 是否被取到为条件，那么被取出的元素的平均位置能表示为

$$\begin{aligned}E[位置]&=pE[X]+(1-p)\frac{E[1+2+\cdots+n-X]}{n-1}\\&=\left(p-\frac{1-p}{n-1}\right)E[X]+\frac{(1-p)n(n+1)}{2(n-1)}\end{aligned}$$

于是，若 $p\geqslant 1/n$，则使 $E[X]$ 达最小而得到最小的平均位置，而若 $p\leqslant 1/n$，则使 $E[X]$ 达到最大而得到最小的平均位置。由于 $E[X]=\sum\limits_{i=0}^{n}P\{X>i\}$，所以由命题 4.6.3 定理得证。

4.7 时间可逆的马尔可夫链

一个不可约正常返马尔可夫链是平稳的，若初始状态的选取

服从平稳分布。（在遍历链的情形这等价于设想此过程从 $t=-\infty$ 开始。）我们说这样一个链处在稳态之中。

现在考虑一个具有转移概率 P_{ij} 及平稳分布 π_i 的平稳马尔可夫链，且假设从某个时刻开始我们把时间倒过来追踪状态的序列。也就是从时刻 n 开始考虑状态序列 X_n，X_{n-1}，…，此状态序列本身也是一个马尔可夫链，转移概率 P_{ij}^* 为

$$
\begin{aligned}
P_{ij}^* &= P\{X_m=j \mid X_{m+1}=i,\ X_{m+2}=i_2,\ \cdots,\ X_{m+k}=i_k\} \\
&= \frac{P\{X_m=j,\ X_{m+1}=i,\ X_{m+2}=i_2,\ \cdots,\ X_{m+k}=i_k\}}{P\{X_{m+1}=i,\ X_{m+2}=i_2,\ \cdots,\ X_{m+k}=i_k\}} \\
&= \frac{P\{X_m=j\}\ P\{X_{m+1}=i \mid X_m=j\}\ P\{X_{m-2}=i_2,\ \cdots,\ X_{m+k}=j_k \mid X_m=j,\ X_{m+1}=i\}}{P\{X_{m+1}=i\}\ P\{X_{m+2}=i_2,\ \cdots,\ X_{m+k}=i_k \mid X_{m+1}=i\}} \\
&= \frac{\pi_j P_{ji} P\{X_{m+2}=i_2,\ \cdots,\ X_{m+k}=j_k \mid X_{m+1}=i\}}{\pi_i P\{X_{m+2}=i_2,\ \cdots,\ X_{m+k}=i_k \mid X_{m+1}=i\}} \\
&= \frac{\pi_j}{\pi_i} P_{ji}
\end{aligned}
$$

于是此逆过程也是一个马尔可夫链，转移概率为

$$P_{ij}^* = \frac{\pi_j P_{ji}}{\pi_i}$$

如果对一切 i，j 有 $P_{ij}^*=P_{ij}$，则称此马尔可夫链是时间可逆的。时间可逆性的条件即是对一切 i，j

$$\pi_i P_{ij} = \pi_j P_{ji} \tag{4.7.1}$$

它可解释为对一切状态 i 与 j，过程从 i 到 j 的比率（即 $\pi_j P_{ij}$）等于它从 j 到 i 的比率（即 $\pi_i P_{ji}$）。应该指出，这是时间可逆性的一个显然的必要条件，因为逆时间方向的从 i 转移到 j 等价于顺时间方向从 j 转移到 i；即，若 $X_m=i$ 及 $X_{m-1}=j$，则逆时间方向看，这是从 i 转移到 j，而顺时间看这就是从 j 转移到 i。

如果能找到总和为 1 的非负数满足（4.7.1），则此马尔可夫链是时间可逆的且这些数表示平稳分布，因为如果

$$x_i P_{ij} = x_j P_{ji}, \qquad \text{对一切 } i,\ j, \quad \sum_{i=0}^{\infty} x_i = 1, \tag{4.7.2}$$

则对 i 求和得

$$\sum_{i=0}^{\infty} x_i P_{ij} = x_j \sum_{i=0}^{\infty} P_{ji} = x_j, \qquad \sum_{i=0}^{\infty} x_i = 1$$

由于平稳分布 π_i 是上式的唯一解，故得对一切 i 有 $x_i=\pi_i$。

例 4.7(a)　一个遍历随机游动。 无需任何计算我们能够证明满足 $P_{i,i+1}+P_{i,i-1}=1$ 的一个遍历链是时间可逆的。注意到在任何时候从 i 到 $i+1$ 的转移次数与以 $i+1$ 到 i 的转移次数相差必不超过 1 就可得出此结论。而后者是因为任何两次从 i 到 $i+1$ 的转移之间必有一次是从 $i+1$ 到 i 的转移（反之亦然），以更高的状态重返状态 i 必须经过状态 $i+1$，因此从 i 到 $i+1$ 转移的比率等于从 $i+1$ 到 i 的比率，从而过程是时间可逆的。

如果我们试着对任意的马尔可夫链求解方程(4.7.2)，那么通常解不存在。例如，由(4.7.2)

$$x_iP_{ij}=x_jP_{ji}$$
$$x_kP_{kj}=x_jP_{jk}$$

（若 $P_{ij}P_{jk}>0$）这包含了

$$\frac{x_i}{x_k}=\frac{P_{ji}P_{kj}}{P_{ij}P_{jk}}$$

一般它不一定等于 P_{ki}/P_{ik}，于是可见时间可逆性的一个必要条件是对一切 i，j，k

(4.7.3)
$$P_{ik}P_{kj}P_{ji}=P_{ij}P_{jk}P_{ki}$$

它等价于说，从状态 i 出发的路径 $i\to k\to j\to i$ 与逆向路径 $i\to j\to k\to i$ 具有相同的概率。为了明瞭为何这是必要的，注意时间可逆性蕴含着一串从 i 到 k 到 j 到 i 的转移发生的比率必须等于一串从 i 到 j 到 k 到 i 的转移发生的比率（为什么?），从而必有

$$\pi_iP_{ik}P_{kj}P_{ji}=\pi_iP_{ij}P_{jk}P_{ki}$$

由此推得(4.7.3)。

事实上，我们能够证明下列

定理 4.7.1　平稳马尔可夫链是时间可逆的，当且仅当对一切 i，任何一条以 i 出发再回到 i 的路径与其逆向路径有相同的概率，即对一切状态 i，i_1，…，i_k

(4.7.4)
$$P_{i,i_1}P_{i_1,i_2}\cdots P_{i_k,i}=P_{i,i_k}P_{i_k,i_{k-1}}\cdots P_{i_1,i}$$

证明　必要性的证明已经指明。为了证明充分性，固定状态 i 与 j 且改写(4.7.4)为

$$P_{i,i_1}P_{i_1,i_2}\cdots P_{i_k,j}P_{ji}=P_{ij}P_{j,i_k}\cdots P_{i_1,i}$$

就一切状态i_1，i_2，…，i_k求和得

$$P_{ij}^{k+1}P_{ji}=P_{ij}P_{ji}^{k+1}$$

因此，

$$P_{ji}\frac{\sum_{k=1}^{n}P_{ij}^{k+1}}{n}=P_{ij}\frac{\sum_{k=1}^{n}P_{ji}^{k+1}}{n}$$

令$n\to\infty$得

$$P_{ji}\pi_j=P_{ij}\pi_i$$

结论得证。

例4.7(b)　名册问题。设给定n个元素(从1到n编号)，它们被排列成有次序的名册，每次要求取出其中一个元素，取到元素i(与过去独立)的概率为P_i，取出之后再放回，但不必放在同一位置。事实上，我们假设取出的元素被移动到名册的前面一位上；例如，现在名册次序为1，3，4，2，5而2被取出，然后新的次序成为1，3，2，4，5。

对于任给的概率向量$\underline{P}=(P_1,P_2,\cdots,P_n)$，能为上述情形建立一个$n!$个状态的马尔可夫链的模型，在任一时刻的状态就是此时的名册次序。用定理4.7.1容易证明这个链是时间可逆的。例如，假设$n=3$，考虑下列从状态(1,2,3)到自身的路径

$$(1,2,3)\to(2,1,3)\to(2,3,1)\to(3,2,1)\to(3,1,2)\to(1,3,2)\to(1,2,3)$$

按正方向及按逆方向的转移概率的乘积都等于$P_1^2P_2^2P_3^2$。由于类似的结果一般情形也成立，所以此马尔可夫链是时间可逆的。

事实上，注意到下列事实也能验证时间可逆性与极限概率。对任意的置换$(i_1,i_2,\cdots i_n)$，由下式给出的概率

$$\pi(i_1,i_2,\cdots i_n)=CP_{i_1}^{n}P_{i_2}^{n-1}\cdots P_{i_n}$$

满足方程式(4.7.1)，其中C取得使

$$\sum_{(i_1,\cdots i_n)}\pi(i_1,\cdots,i_n)=1$$

因此，我们得到了链是可逆的第二种证明，且平稳分布如上给出。

逆向链的概念甚至在过程不是时间可逆时也是有用的。为了说明这点我们从下列定理开始。

定理4.7.2　考虑一个转移概率为P_{ij}的不可约马尔可夫链，

如果能找到非负数 π_i，$i\geqslant 0$，其和为 1，及一个转移矩阵 $P^*=[P_{ij}^*]$ 使得

$$\pi_i P_{ij}=\pi_j P_{ji}^* \tag{4.7.6}$$

则 π_i，$i\geqslant 0$，是平稳概率分布，P_{ij}^* 是逆向链的转移概率。

证明 将上列等式对一切 i 求和得

$$\sum_{i=0}^{\infty}\pi_i P_{ij}=\pi_j\sum_{i=0}^{\infty}P_{ji}^*=\pi_j$$

因此 π_i 是正向链的平稳分布（而且也是逆向链的平稳分布；为什么?）。因为

$$P_{ji}^*=\frac{\pi_i P_{ij}}{\pi_j}$$

故得 P_{ij}^* 是逆向链的转移概率。

定理 4.7.2 的重要性在于有时我们能猜测逆向链的本性，而后利用方程(4.7.6)同时得到平稳分布与 P_{ij}^*。

例 4.7(c) 让我们再考虑例 4.3(c)，它处理离散时间更新过程的年龄，以 X_n 记一个更新过程在时刻 n 的年龄，其来到间隔都是整数。因此此马尔可夫链的状态总是增加 1，直到它到达一个按来到间隔的分布选定的值，而后降为 1。所以逆向过程总是减少 1，直到它到达状态 1，此时它跳到一个按来到间隔的分布选定的状态，因此看来逆向过程正是过剩或剩余寿命过程。

以 P_i 记来到间隔等于 i 的概率，$i\geqslant 1$，看来很可能有

$$P_{1i}^*=P_i,\quad P_{i,i-1}^*=1,\qquad i>1$$

因为

$$P_{i1}=\frac{P_i}{\sum\limits_{i=j}^{\infty}P_j}=1-P_{i,i+1},\qquad i\geqslant 1$$

对上面给出的逆向链，由(4.7.6)我们需要有

$$\frac{\pi_i P_i}{\sum\limits_{j=i}^{\infty}P_j}=\pi_1 P_i$$

或

$$\pi_i=\pi_1 P\{X\geqslant i\}$$

其中 X 是一个来到间隔，因为 $\sum\limits_{i=1}^{\infty}\pi_i=1$，所以上式必定要有

$$1=\pi_1\sum_{i=1}^{\infty}P\{X\geqslant i\}$$
$$=\pi_1E[X],$$

从而为了使逆向链如同我们所猜测的那样，必须有

(4.7.7) $$\pi_i=\frac{P\{X\geqslant i\}}{E[X]}$$

为了完成逆向过程是剩余寿命及极限概率为(4.7.7)所给定的证明，我们需要验证：

$$\pi_iP_{i,i+1}=\pi_{i+1}P^*_{i+1,i}$$

或等价地

$$P\{X\geqslant i\}[1-\frac{P_i}{P\{X\geqslant i\}}]=P\{X\geqslant i+1\}$$

而此式可立即得到。

因此，察看逆向链我们能够证明它就是更新过程的剩余寿命，且同时得到(既是剩余寿命又是年龄的)极限分布。事实上，这个例子补充阐明了为什么更新过程的剩余寿命与年龄有同样的极限分布。

利用逆向链得到极限概率的技巧在第五章处理连续时间马尔可夫链时将进一步运用。

4.8 半马尔可夫过程

一个半马尔可夫过程是一个随机过程，其状态变化遵循一个马尔可夫链，而状态变化的时间间隔是随机变量。更确切地说，考虑一具有状态 0，1，…，的随机过程，满足以下条件：每当它进入状态 i，$i\geqslant 0$ 时

(i) 下一个进入的状态是 j 的概率为 P_{ij}，$i,j\geqslant 0$。

(ii) 在下一个进入的状态是 j 的条件下，直到发生从 i 到 j 转移为止的时间有分布 F_{ij}。

若以 $Z(t)$ 记时刻 t 的状态，则 $\{Z(t),t\geqslant 0\}$ 称为半马尔可夫过程。

一个半马尔可夫过程不具有给定现在的状态时将来与过去独

立的马尔可夫性。为了预测将来，我们不仅要知道现在的状态，还要知道在此状态已停留了多少时间。当然，在转移的时刻，我们所需要知道的只是新的状态（而无需关于过去的情况），一个马尔可夫链是一个半马尔可夫过程，其

$$F_{ij}(t)=\begin{cases}0, & t<1\\ 1, & t\geqslant 1\end{cases}$$

即一个马尔可夫链的转移时间恒为 1。

以 H_i 记半马尔可夫过程在转移之前处于状态 i 的时间的分布。对下一个状态取条件可见

$$H_i(t)=\sum_j P_{ij}F_{ij}(t)$$

以 μ_i 记其均值，即

$$\mu_i=\int_0^\infty x dH_i(x)$$

若以 X_n 记第 n 个到达的状态，则 $\{X_n, n\geqslant 0\}$ 是一个转移概率为 P_{ij} 的马尔可夫链，它称为半马尔可夫过程的嵌入马尔可夫链。若此嵌入马尔可夫链是不可约的，则称此半马尔可夫过程是不可约的。

以 T_{ii} 记相继进入状态 i 之间的时间，且令 $\mu_{ii}=E[T_{ii}]$。运用交错更新过程理论，导出半马尔可夫过程的极限概率的表达式是一件简单的事情。

命题 4.8.1

若半马尔可夫过程是不可约的，且 T_{ii} 具有非格点的分布与有限的均值，则

$$P_i\equiv\lim_{t\to\infty}P\{Z(t)=i\mid Z(0)=j\}$$

存在且与初始状态无关。更进一步，

$$P_i=\frac{\mu_i}{\mu_{ii}}$$

证明 每当过程进入状态 i 就说一个循环开始，且当过程处于状态 i 时就说过程是“开的”，不在状态 i 时则它是“关的”。于是我们有了一个

（当 $Z(0)\neq i$ 时是延迟的）交错更新过程，它开着的时间有分布 H_i，而它的循环时间为 T_{ii}。因此，由第三章的命题 3.4.4 结论得证。

作为一个系我们注意到 P_i 也等于长时间之后过程处于状态 i 的时间的比率。

系 4.8.2

若半马尔可夫过程是不可约的，且 $\mu_{ii}<\infty$，则以概率 1

$$\frac{\mu_i}{\mu_{ii}}=\lim_{t\to\infty}\frac{[0,\ t]\ \text{中处于状态}\ i\ \text{的时间}}{t}$$

即 μ_i/μ_{ii} 等于长时期之后处于状态 i 的时间的比率。

证明 由第三章的命题 3.7.2 可得。

虽然命题 4.8.1 给出了极限概率的表达式，但它并不是一个实际计算 P_i 的方法。为要计算 P_i，假设嵌入马尔可夫链 $\{X_n,n\geqslant 0\}$ 不可约正常返，又设它的平稳分布是 π_j，$j\geqslant 0$。即 π_j，$j\geqslant 0$，是

$$\pi_j=\sum_{i=0}^{\infty}\pi_i P_{ij}$$

$$\sum_{j=0}^{\infty}\pi_j=1$$

的唯一解，且 π_j 可解释为等于 j 的 X_n 所占的比率。（若马尔可夫链是非周期的，则 π_j 也等于 $\lim\limits_{n\to\infty}P\{X_n=j\}$。）既然 π_j 等于进入状态 j 的转移的比率，而 μ_j 是每一次转移到 j 后处于状态 j 的平均时间，所以直观地看来极限概率应与 $\pi_j\mu_j$ 成比例。现在我们证明这一点。

定理 4.8.3 假设命题 4.8.1 的条件成立，且进一步假设嵌入马尔可夫链 $\{X_n,n\geqslant 0\}$ 是正常返的，则

$$P_i=\frac{\pi_i\mu_i}{\sum\limits_j\pi_j\mu_j}$$

证明 定义记号如下：

$Y_i(j)=$第 j 次到达状态 i 后在状态 i 逗留的时间，$i,j\geqslant 0$。

$N_i(m)$ = 在半马尔可夫过程的前 m 次转移中到达状态 i 的次数。

利用上述记号可见，在前 m 次转移中处于状态 i 的时间的比率（记为 $P_{i=m}$）如下：

$$P_{i=m}=\frac{\sum_{j=1}^{N_i(m)} Y_i(j)}{\sum_{i=0}^{\infty}\sum_{j=1}^{N_i(m)} Y_i(j)} \tag{4.8.1}$$

$$=\frac{\frac{N_i(m)}{m}\sum_{j=1}^{N_i(m)}\frac{Y_i(j)}{N_i(m)}}{\sum_{i=0}^{\infty}\frac{N_i(m)}{m}\sum_{j=1}^{N_i(m)}\frac{Y_i(j)}{N_i(m)}}$$

由于 $m\to\infty$ 时，$N_i(m)\to\infty$，从强大数定律推得

$$\sum_{j=1}^{N_i(m)}\frac{Y_i(j)}{N_i(m)}\to\mu_i$$

由更新过程的强极限律得

$$\frac{N_i(m)}{m}\to(E[\text{两次到达 } i \text{ 之间的转移次数}])^{-1}=\pi_i$$

因此，在(4.8.1)中令 $m\to\infty$ 证明了

$$\lim_{m\to\infty}P_{i=m}=\frac{\pi_i\mu_i}{\sum_j\pi_j\mu_j}$$

证毕。

从定理 4.8.3 得出，极限概率只依赖于转移概率 P_{ij} 与平均时间 μ_i，$i,j\geqslant 0$。

例 4.8(a) 一部机器可能有三种状态：良好，尚好，或损坏。假设良好状态的机器保持良好的平均时间为 μ_1，然后分别以概率 3/4 与 1/4 转到尚好或损坏的状态，处于尚好状态的机器保持此状态的平均时间为 μ_2，然后损坏，一部损坏的机器将被修理，修理所用的平均时间为 μ_3，修理好的机器以概率 2/3 处于良好状态，以概率 1/3 处于尚好状态。机器处于各状态的时间的比率是多少？

解 设状态是 1，2，3，π_i 满足

$$\pi_1+\pi_2+\pi_3=1$$

$$\pi_1=\frac{2}{3}\pi_3$$

$$\pi_2=\frac{3}{4}\pi_1+\frac{1}{3}\pi_3$$

$$\pi_3=\frac{1}{4}\pi_1+\pi_2$$

其解是

$$\pi_1=\frac{4}{15},\qquad \pi_2=\frac{1}{3},\qquad \pi_3=\frac{2}{5}$$

因此，机器处于状态 i 的时间的比率 P_i 为

$$P_1=\frac{4\mu_1}{4\mu_1+5\mu_2+6\mu_3}$$

$$P_2=\frac{5\mu_2}{4\mu_1+5\mu_2+6\mu_3}$$

$$P_3=\frac{6\mu_3}{4\mu_1+5\mu_2+6\mu_3}$$

确定半马尔可夫过程的极限分布的问题并未因为求出了 P_i 而完全解决。例如，我们可能想求在时刻 t 处于状态 i，下一次转移在时刻 $t+x$ 之前并转移到状态 j 的概率在 $t\to\infty$ 时的极限。为了表示这个概率令

$Y(t)=$从 t 到下一次转移的时间

$S(t)=t$ 之后首次转移进入的状态

为了计算

$$\lim_{t\to\infty}P\{Z(t)=i,Y(t)>x,S(t)=j\}$$

我们将再次使用交错更新过程理论。

定理 4.8.4 若半马尔可夫过程不可约，且是非格点的，则

$$\lim_{t\to\infty}P\{Z(t)=i,Y(t)>x,S(t)=j\mid Z(0)=k\}=\frac{P_{ij}\int_x^\infty\overline{F}_{ij}(y)dy}{\mu_{ii}}\tag{4.8.2}$$

证明 每当过程进入状态 i，就说一个循环开始了；若状态为 i，在状态 i 至少停留 x 时间且下一个状态是 j，就说它是“开的”。其余情况说它是

"关的"，这样，我们就有一个交错更新过程。对在 i 之后的状态是否是 j 取条件，可见

$$E[\text{一个循环中的“开”时}] = P_{ij}E[(X_{ij}-x)^+]$$

其中 X_{ij} 是具有分布 F_{ij} 的随机变量，它代表作出从 i 到 j 的转移的时间，而 $y^+=\max(0,y)$。因此

$$\begin{aligned}E[\text{循环中的“开”时}] &= P_{ij}\int_0^\infty P\{X_{ij}-x>a\}da \\ &= P_{ij}\int_0^\infty \overline{F}_{ij}(a+x)da \\ &= P_{ij}\int_x^\infty \overline{F}_{ij}(y)dy\end{aligned}$$

因为 $E[\text{循环时间}]=\mu_{ii}$，由交错更新过程理论保证定理之结论。

用同样的技巧(或将(4.8.2)对 j 求和)我们可以证明下列

系 4.8.5

若半马尔可夫过程不可约且是非格点的，则

(4.8.3) $$\lim_{t\to\infty}P\{Z(t)=i,Y(t)>x|Z(0)=k\} = \int_x^\infty \overline{H}_i(y)dy/\mu_{ii}$$

注记

(1) 当然，定理 4.8.4 与系 4.8.5 中的极限概率也可解释为长时间之后的比率。例如，考虑这样的时刻，其时半马尔可夫过程处于状态 i，之后的 x 时间内无转移，而后将转移到状态 j，定理 4.8.4 给出了长时间之后这种时刻所占的比率。

(2) 将(4.8.3)乘以及除以 μ_i，且利用 $P_i=\mu_i/\mu_{ii}$，得

$$\lim_{t\to\infty}P\{Z(t)=i,Y(t)>x\}=P_i\overline{H}_{i,e}(x)$$

其中 $H_{i,e}$ 是 H_i 的平衡分布。因此处于状态 i 的极限概率是 P_i，且已知 t 时的状态是 i，到发生转移为止的时间(当 t 趋于 ∞)服从 H_i 的平衡分布。

习 题

1. 一家商店使用如下的 (s,S) 定货策略贮备某种商品；如果在一天的开始其

供应量是 x，则它定购

$$\begin{cases} 0, & \text{当 } x \geqslant s \\ S-x, & \text{当 } x < s \end{cases}$$

定货立即被满足，每天的需求量是独立的，且以概率 α_j 取值 j，不能立即满足的一切需求皆消失。令 X_n 表示在第 n 天结束时的存货水平，论证 $\{X_n, n\geqslant 1\}$ 是一个马尔可夫链，且计算其转移概率。

2. 对一个马尔可夫链证明：

$$P\{X_n=j \mid X_{n_1}=i_1, \cdots, X_{n_k}=i_k\}=P\{X_n=j \mid X_{n_k}=i_k\}$$

当 $n_1<n_2<\cdots<n_k<n$ 时皆成立。

3. 证明：如果状态的个数是 n，且如果状态 j 可从状态 i 到达，则它可用 n 步或更少的步数到达。

4. 证明：

$$P_{ij}^n=\sum_{k=0}^{n} f_{ij}^k P_{jj}^{n-k}$$

5. 证明：对称随机游动在二维时是常返的，而在三维时是滑过的。

6. 对从 0 出发的对称随机游动：

(a) 回到 0 的平均时间是多少？

(b) 以 N_n 记到时刻 n 为止返回的次数。证明：

$$E[N_{2n}]=(2n+1)\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}-1$$

(c) 用 (b) 及斯特林近似公式证明：对很大的 n，$E[N_n]$ 与 $\sqrt{n}$ 成比例。

7. 设 $X_1, X_2, \cdots$ 是独立的随机变量，使得 $P\{X_i=j\}=\alpha_j$，$j\geqslant 0$。如果 $X_n>\max(X_1, \cdots, X_{n-1})$，其中 $X_0=-\infty$，就说在时刻 n 产生了一个记录。若在时刻 n 产生了一个记录，称 X_n 为记录值，以 R_i 记第 i 个记录值。

(a) 论证：$\{R_i, i\geqslant 1\}$ 是一个马尔可夫链，并计算其转移概率。

(b) 以 T_i 记第 i 个与第 $i+1$ 个记录之间的时间。$\{T_i, i\geqslant 1\}$ 是一个马尔可夫链吗？$\{(R_i, T_i), i\geqslant 1\}$ 呢？是马尔可夫链时，计算转移概率。

8. 如果 $f_{ii}<1$ 及 $f_{jj}<1$，证明：

(a) $\sum_{n=1}^{\infty} P_{ij}^n<\infty$　　　(b) $f_{ij}=\dfrac{\sum_{n=1}^{\infty} P_{ij}^n}{1+\sum_{n=1}^{\infty} P_{jj}^n}$

9. 转移概率矩阵 P 称为双随机的，若对一切 j

$$\sum_{i=0}^{\infty} P_{ij}=1$$

即各列之和都等于 1。若一个双随机的链有 n 个状态且是遍历的，计算其极限概率。

10. 证明：正常返与零常返是类性质。

11. 证明：在一个有限马尔可夫链中无零常返状态，且不可能所有的状态都是非常返的。

12. 在 $M/G/1$ 系统中(例 4.3(a))假设 $\rho<1$ 从而平稳分布存在，计算 $\pi'(s)$，且令 $s\to1$ 取极限求 $\sum_{i=0}^{\infty} i\pi_i$。

13. 一个人有 r 把伞用于上下班，如果一天的开头(结束)他是在家(办公室)中而且天下雨，只要有伞可取到，他就拿一把到办公室(家)中。如果天不下雨，那么他绝不带伞，假定一天的开始(结束)下雨的概率为 p，与过去的情况独立。

(a) 定义一个有 $r+1$ 个状态的马尔可夫链，有助于我们确定此人被淋湿的次数的比率。(注意：如果天下雨而全部伞在另一处那么他要被淋湿。)

(b) 计算极限概率。

(c) 占多少比率的次数此人要被淋湿。

14. 考虑一个正常返不可约周期马尔可夫链，且以 π_j 记长时间之后处于状态 j，$j\geqslant0$，的时间的比率。证明：π_j，$j\geqslant0$，满足 $\pi_j=\sum_{i=0}^{\infty}\pi_i P_{ij}$，$\sum_{j=0}^{\infty}\pi_j=1$。

15. 一个马尔可夫链有一个吸收状态 0 (即 $P_{00}=1$) 与滑过状态 1，2，…，设 T_i 是链从 i 出发到达状态 0 的时间，且令 $M_i=E[T_i]$，$i>0$，

(a) 证明：

$$M_i=1+\sum_{j>0} P_{ij}M_j$$

(b)令 $\sigma_i(n)=P\{T_i>n\}$。导出用 $\sigma_j(n)$，$j>0$，表示 $\sigma_i(n+1)$ 的一个公式。

16. 考虑一个具有状态 0，1，2，…的马尔可夫链，其转移概率：

$$P_{i,i+1}=p_i=1-P_{i,i-1}$$

其中 $p_0=1$。找出为使链正常返诸 p_i 满足的充要条件，且就这种情况计算其极限概率。

17. 在赌徒输光问题中，从 i 元开始直到赌徒到达 0 或 N 元为止，计算赌局的平均次数。

18. 在赌徒输光问题中证明：

$$P\{\text{他在下一局赌博中赢}\mid\text{目前的赌金为 } i\text{，他最终达到 } N \text{ 元}\}$$

$$=\begin{cases} p\,[1-(q/p)^{i+1}]\,/\,[1-(q/p)^{i}], & \text{若 } p\neq \frac{1}{2} \\ (i+1)\,/2i, & \text{若 } p=\frac{1}{2} \end{cases}$$

19. 每天 n 个元素中的一个被需求；第 i 个被需求的概率为 P_i，$i\geqslant 1$，$\sum_{i=1}^{n} P_i=1$。这些元素总是被排成有次序的名册，且它的变动如下：被挑选到的元素移到名册的最前面，而所有其它元素的相对位置不变。定义在任意时刻的状态是该时刻名册的次序。

(a) 论证上述过程是一个马尔可夫链。

(b) 对任意的状态 $i_1,\cdots,i_n$（它是 $1,2,\cdots,n$ 的一个置换），以 $\pi(i_1,\cdots,i_n)$ 记极限概率。证明：

$$\pi(i_1,\cdots,i_n)=P_{i_1}\frac{P_{i_2}}{1-P_{i_1}}\cdots\frac{P_{i_{n-1}}}{1-P_{i_1}-P_{i_2}-\cdots-P_{i_{n-2}}}$$

20. 假设两个独立序列 $X_1,X_2,\cdots$ 与 $Y_1,Y_2,\cdots$ 来自某个实验室，它们代表着成功概率 P_1 与 P_2 未知的贝努里试验。即 $P\{X_i=1\}=1-P\{X_i=0\}=P_1$，$P\{Y_i=1\}=1-P\{Y_i=0\}=P_2$，且所有的随机变量是独立的。为了确定 $P_1>P_2$ 还是 $P_2>P_1$，我们用下列检验法。选取某个正整数 M，且在 N 停止，N 是使得

$$X_1+\cdots+X_n-(Y_1+\cdots+Y_n)=M$$

或

$$X_1+\cdots+X_n-(Y_1+\cdots+Y_n)=-M$$

的第一个 n 的值。在前一情形我们就断定 $P_1>P_2$，而在后一种情形就断定 $P_2>P_1$。证明：当 $P_1>P_2$ 时，犯错误的概率（即 $P_2>P_1$）是

$$P\{\text{错误}\}=\frac{1}{1+\lambda^M}$$

而被观察的数对的个数之平均值是

$$E[N]=\frac{M(\lambda^M-1)}{(P_1-P_2)(\lambda^M+1)}$$

其中

$$\lambda=\frac{P_1(1-P_2)}{P_2(1-P_1)}$$

（提示：将问题与赌徒输光问题联系起来）。

21. 捕捉苍蝇的一只蜘蛛依循一马尔可夫链在位置 1 与 2 之间移动，其起始位置为 1，转移矩阵为$\begin{pmatrix}0.7 & 0.3\\ 0.3 & 0.7\end{pmatrix}$，未知觉到蜘蛛的苍蝇的初始位置为 2，并依照转移矩阵为$\begin{pmatrix}0.4 & 0.6\\ 0.6 & 0.4\end{pmatrix}$的马尔可夫链移动，只要它们在同一位置相遇，蜘蛛就捉住苍蝇从而结束捕捉。

证明：捕捉的过程，除非知道它结束时的位置，可用三个状态的马尔可夫链来描述，其中一个是吸收状态代表捕捉结束，而另外的两个代表蜘蛛与苍蝇处在不同的位置，对此链求转移概率矩阵。

求在时刻 n 蜘蛛与苍蝇都处于各自的初始位置的概率。

捕捉过程的平均持续时间是多少？

22. 考虑整数点上的一个简单随机游动，其中一质点每一步朝正方向移动一步的概率为 p，朝负方向移动一步的概率也为 p，而以概率 $q=1-2p$ $(0<p<\frac{1}{2})$ 留在原处，假设在原点放置一个吸收壁（即 $P_{00}=1$），而在 N 处放一反射壁（即 $P_{N,N-1}=1$），又质点起始于 n $(0<n<N)$。

证明：质点被吸收的概率为 1。且求被吸收所需的平均步数。

23. 给定$\{X_n,n\geqslant 0\}$为一分支过程：

(a) 论证：X_n 或收敛于 0 或趋于无穷。

(b) 证明：

$$\operatorname{Var}(X_n|X_0=1)=\begin{cases}\sigma^2\mu^{n-1}\dfrac{\mu^n-1}{\mu-1}, & 若\ \mu\neq 1\\ n\sigma^2, & 若\ \mu=1\end{cases}$$

其中 μ 与 σ^2 是一个个体的后代个数的均值和方差。

24. 在一分支过程中每个个体的后代个数具有二项分布，其参数是 2，p。从一个个体开始，计算：

(a) 灭绝的概率；

(b) 到第三代群体首次灭绝的概率。

假设开始时不是一个个体，初始的群体总数 Z_0 是一随机变量，服从均值为 λ 的泊松分布。证明：此时，对 $p>1/2$，灭绝概率为

$$\exp\{\lambda(1-2p)/p^2\}$$

25. 对于 4.6.1 节中的马尔可夫链模型，即

$$P_{ij}=\frac{1}{i-1},\quad j=1,\cdots,i-1,\quad i>1$$

假设初始状态是 $N\equiv\binom{n}{m}$，其中 $n>m$。证明：当 n，m 与 $n-m$ 很大时从状态 N 到达 1 的步数渐近地具有泊松分布，其均值为

$$m\left[c\log\frac{c}{c-1}+\log(c-1)\right]$$

其中 $c=n/m$。（提示：用斯特林近似公式。）

26. 对于任意的无穷序列 x_1，x_2，…，每当序列增减的方向变换，我们就说一个新的长游程开始了。譬如，若序列开头为 5，2，4，5，6，9，3，4，那么其中有三个长游程——即 (5，2)，(4，5，6，9)，与(3，4)。设 X_1，X_2，…是独立的(0,1)均匀分布随机变量，以 I_n 记第 n 个长游程的初始值，论证：$\{I_n, n\geqslant 1\}$ 是一具有连续状态空间的马尔可夫链，其转移概率密度为：

$$p(y|x)=e^{1-x}+e^{x}-e^{|y-x|}-1$$

27. 考虑在例 4.7(b)中给出的名册模型。在移前一位的规则下，利用时间可逆性证明：元素 j 在 i 之前的极限概率（记为 $P\{j$ 在 i 之前$\}$）满足

$$P\{j\text{ 在 }i\text{ 之前}\}>\frac{P_j}{P_j+P_i}\qquad \text{当 }P_j>P_i\text{ 时}$$

28. 设有 n 个位置，对全部$\binom{n}{2}$对位置的每一对，给定了一个正数 $d_{ij}=d_{ji}$。一质点依照如下规则移动位置：在任一时刻，若此质点在位置 i 上，那么下一步它将移动到位置 j 的概率 P_{ij} 为：

$$P_{ij}=\frac{d_{ij}}{\sum\limits_{j\neq i}d_{ij}}$$

证明：位置的序列是时间可逆的马尔可夫链，并求极限概率。

29. 考虑一个具有转移概率 P_{ij} 及极限概率 π_i 的时间可逆马尔可夫链；现在考虑状态截到 0，1，…，M 为止的链；也即，截尾链的转移概率 $\overline{P}_{ij}$ 是

$$\overline{P}_{ij}=\begin{cases}P_{ij}\Big/\sum\limits_{j=0}^{M}P_{ij}, & 0\leqslant i,\ j\leqslant M\\ 0, & \text{其它}\end{cases}$$

证明：此截尾链也是时间可逆的且其极限概率为

$$\bar{\pi}_i=\frac{\pi_i\sum_{j=0}^{M}P_{ij}}{\sum_{i=0}^{M}\pi_i\sum_{j=0}^{M}P_{ij}}$$

30. 证明：若一个有限状态的遍历马尔可夫链满足对一切 $i\neq j$，$P_{ij}>0$，则它是时间可逆的当且仅当对于一切 i，j，k 有

$$P_{ij}P_{jk}P_{ki}=P_{ik}P_{kj}P_{ji}$$

31. 以 $\{X_n,n\geqslant 1\}$ 记具有可数状态空间的一个不可约马尔可夫链。今考虑一新的随机过程 $\{Y_n,n\geqslant 0\}$，它只接纳原马尔可夫链的第 n 个在 0 与 N 之间的值。例如，若 $N=3$，且 $X_1=1$，$X_2=3$，$X_3=5$，$X_4=6$，$X_5=2$，则 $Y_1=1$，$Y_2=3$，$Y_3=2$。

(a) $\{Y_n,n\geqslant 0\}$ 是一个马尔可夫链吗？简要地解释。

(b) 以 π_j 记 $\{X_n,n\geqslant 1\}$ 处于状态 j 的时间的比率。如果对一切 j，$\pi_j>0$，问 $\{Y_n,n\geqslant 0\}$ 处于状态 0，1，$\cdots N$ 中的每一个的时间比率是多少？

(c) 假设 $\{X_n\}$ 是零常返的，且以 $\pi_i(N)$，$i=0,1,\cdots,N$，记 $\{Y_n,n\geqslant 0\}$ 长时间之后处于状态 i 的时间的比率。证明：

$\pi_j(N)=\pi_i(N)E[X$ 过程在两次到达 i 之间处于 j 的时间$]$，$j\neq i$。

(d) 用 (c) 去论证，在一个对称随机游动中，在返回到原点之前到达状态 i 的平均次数等于 1。

(e) 如果 $\{X_n,n\geqslant 0\}$ 是时间可逆的，证明 $\{Y_n,n\geqslant 0\}$ 亦然。

32. 开始 M 个球被分布在 m 个罐中。每次随机地从中抽取一球，不论从哪一罐中取出它被随机地放入其余 $m-1$ 个罐中，考虑状态为向量 $(n_1,n_2,\cdots n_m)$ 的马尔可夫链，其中 n_i 表示在罐 i 中的球数。对此马尔可夫链猜测其极限概率，而后验证你的猜测，且同时证明此马尔可夫链是时间可逆的。

33. 对一个遍历的半马尔可夫过程：

(a) 计算此过程从 i 到 j 转移的比率。

(b) 证明：

$$\sum_{i=0}^{\infty}P_{ij}/\mu_{ii}=1/\mu_{jj}$$

(c) 证明：过程处于状态 i 且接下来在时间 x 内转移到 j 的时间的比率为

$P_{ij}\eta_{ij}/\mu_{ii}$，其中 $\eta_{ij}=\int_0^\infty \overline{F}_{ij}(t)dt$。

(d) 证明：处于状态 i 且接下来在时间 x 内转移到 j 的时间的比率为

$$\frac{P_{ij}\eta_{ij}}{\mu_{ii}}F_{i,j}^e(x)$$

其中 F_{ij}^e 是 F_{ij} 的平衡分布。

34. 对遍历半马尔可夫过程，推导出：已知 $X(t)=i$ 时，在 t 之后到达的下一个状态是 j 的条件概率，在 $t\to\infty$ 时的表达式。

35. 一出租汽车流动于三个位置之间。当它到达位置 1 时然后等可能地去 2 或 3。当它到达 2 时，接着它将以概率 1/3 到 1 而以概率 2/3 到 3。从 3 总是开往 1。在位置 i 与 j 之间的平均时间是 $t_{12}=20$，$t_{13}=30$，$t_{23}=30$ $(t_{ij}=t_{ji})$。

此出租汽车最近停的位置为 i 的(极限)概率是多少？$i=1$，2，3。

此出租汽车朝位置 2 开的(极限)概率是多少?

有多少比例的时间此出租汽车是从 2 开到 3?注意:此出租汽车一到达一个位置立即就又开出。

参考文献

简单（不一定对称）随机游动的绝对值是一马尔可夫链首先由雪希纳所指出。算法的有效性的马尔可夫链模型(4.6.1 节)取自文献 11。那里证明了若 $N=\binom{n}{m}$（此 n 代表变量的个数而 m 代表线性规划中的约束的个数)，则所需的迭代次数服从泊松分布，均值为 $\log N\sim m[c\log(c/c-1)+\log\ (c-1)]$，其中 $c=n/m$。4.6.2 节的模型是众所周知的（例如，见文献 8)；然而，我们直观地得出极限密度而后验证它满足平稳分布方程的方法比文献 8 中所述更为直接了当。4.6.3 节关于名册排序问题的结果来自文献 6。他们证明了，移前一位规则在那些把取出的元素往前移而保留其余元素相对位置不变的规则中，对某一类 p 向量是最优的。最近有一个反例(见文献 1) 表明移前一位规则在一切可能的规则之中不一定是最优的。

参考文献 2、3、5、7、9、10 与 13 给出了马尔可夫链的其它处理方法。文献 2 与 13 也有关于半马尔可夫过程的章节，文献 4 是分支过程的标准教材。

1. E. Anderson, P. Nash, and R. Weber, "A Counterexample to a Conjecture on Optimal List Ordering," *Journal of Applied Probability* (to appear).
2. E. Cinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, N. J. 1975.
3. D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes*, Methuen, London, 1965.
4. T. Harris, *The Theory of Branching Processes*, Springer-Verlag, Berlin, 1963.
5. D. Isaacson and R. Madsen, *Markov Chains Theory and Applications*, Wiley, New York, 1976.
6. Y. C. Kan and S. Ross, "Optimal List Order Under Partial Memory Constraint," *Journal of Applied Probability*, 17 (1980), pp. 1004—1015.
7. J. Kemeny, L. Snell, and A. Knapp, *Denumerable Markov Chains*, Van Nostrand, 1966.
8. D. Knuth, *Sorting and Searching*, *Volume* 3 *of The Art of Computer Programming*, Addison-Wesley, Reading, Mass. 1973.
9. E. Parzen, *Stochastic Processes*, Holden-Day, San Francisco, 1962. (伊・帕尔逊，随机过程，邓永录、杨振明译，高等教育出版社，1987.)
10. N. U. Prabhu, *Stochastic Processes*, Macmillan, New York, 1965.
11. S. Ross, "A Simple Heuristic Approach to Simplex Efficiency," Operations Research Center Report, University of California, Berkeley. 1981.
12. S. Ross, *Introduction to Probability Models*, 2nd ed., Academic Press, New York, 1981.
13. S. Ross, *Applied Probability Models With Optimization Applications*, *Holden-Day*, *San Francisco*, 1970.

5
连续时间马尔可夫链

5.1 引言

本章中我们考虑与离散时间马尔可夫链类似的连续时间马尔可夫链。如离散情形一样，它们由马尔可夫性刻划，即已知现在的状态时将来与过去独立。

在 5.2 节中，我们定义连续时间马尔可夫链且把它们与第四章的离散时间马尔可夫链相联系。在 5.3 节中，我们引入一类重要的连续时间马尔可夫链，即所谓生灭过程。这些过程可用作在任何时刻其总量的变化仅为一个单位的群体的模型。在 5.4 节中，我们导出两组描述系统的概率规律的微分方程——向前与向后方程。5.5 节的内容是确定连续时间马尔可夫链的有关的极限(或长时间后的)概率。在 5.6 节中，我们考虑时间可逆的问题。其中，我们证明一切生灭过程是时间可逆的，而后阐明这事实对于排队系统的重要性。在这一节中也提供了时间可逆性对随机群体模型的应用。在 5.7 节中，我们阐明逆向链的重要性，即使过程不是时间可逆的。利用它我们研究排队网络模型，导出爱尔朗消失公式，分析共用加工系统。5.8 节中我们表明如何“一致化”马尔可夫链——对于数值计算有用的一种技巧。

5.2 连续时间马尔可夫链

考虑取非负整数值的连续时间随机过程$\{X(t),t\geqslant 0\}$。与第四章中给出的离散时间马尔可夫链的定义类似，过程$\{X(t),t\geqslant 0\}$称为连续时间马尔可夫链，如果对一切s，$t\geqslant 0$及非负整数i，j，$x(u),0\leqslant u\leqslant s$，有

$$\begin{aligned}&P\{X(t+s)=j|X(s)=i,X(u)=x(u),0\leqslant u<s\}\\&\quad=P\{X(t+s)=j|X(s)=i\}\end{aligned}$$

换言之，连续时间马尔可夫链是具有马尔可夫性的随机过程，即已知现在s时的状态及一切过去的状态的条件下在将来时刻$t+s$的状态的条件分布只依赖现在的状态而与过去独立。若又有$P\{X(t+s)=j|X(s)=i\}$与s无关则称连续时间马尔可夫链具有平稳的或齐次的转移概率。将假定我们所考虑的马尔可夫链都有平稳转移概率。

假设在某时刻，比如说时刻0，马尔可夫链进入状态i，而且假设在接下来的s个单位时间中过程未离开状态i（即未发生转移）。在随后的t个单位时间中过程仍不离开状态i的概率是多少呢？为了回答这个问题，注意到因为在时间s过程处于状态i，从马尔可夫性得在区间$[s,s+t]$中它仍然处于状态i的概率正是它处于状态i至少t个单位时间的（无条件）概率。也即若以τ_i记过程在转移到另一状态之前停留在状态i的时间，则对一切s，$t\geqslant 0$有

$$P\{\tau_i>s+t|\tau_i>s\}=P\{\tau_i>t\}$$

因此，随机变量τ_i是无记忆的必有指数分布。

事实上，上面的讨论给了我们构造连续时间马尔可夫链的一个方法。也即它是一个具有如下性质的随机过程，每当它进入状态i：

(i) 在转移到另一状态之前 处于状态i的时间服从指数分布，

参数为ν_i；

(ii) 当过程离开状态i时，接着以某个概率，譬如P_{ij}，进入状态j，$\sum_{j\neq i}P_{ij}=1$。

$\nu_i=\infty$的状态i称为瞬时状态，因为一旦进入此状态立即就离开。尽管这种状态在理论上是可能的，我们将始终假设对一切i，$0\leqslant\nu_i<\infty$。(如果$\nu_i=0$，则称状态i为吸收的，因为一旦进入这一状态就永不再离开了。) 因此，实际上一个连续时间马尔可夫链是一个这样的随机过程，它按照一个(离散时间)的马尔可夫链从一个状态转移到另一个状态，但在转移到下一状态之前，它在各个状态停留的时间服从指数分布。此外在状态i过程停留的时间与下一个到达的状态必须是独立的随机变量。因为若下一个到达的状态依赖于τ_i，那么过程处于状态i已有多久的信息与下一个状态的预报有关——这就与马尔可夫假定矛盾了。

一个连续时间马尔可夫链称为规则的，若以概率1在任意有限时间内的转移次数是有限的。一个非规则的马尔可夫链的例子是

$$P_{i,i+1}=1,\qquad \nu_i=i^2$$

能够证明这个马尔可夫链总是从状态i到$i+1$，停留在状态i的时间服从均值为$1/i^2$的指数分布，它将以正的概率在任意长为t，($t>0$) 的时间区间内作无限多次转移。然而我们从现在起将假设所考虑的全部马尔可夫链是规则的(在习题中将给出规则性的某些充分条件)。

对一切$i\neq j$，q_{ij}定义为

$$q_{ij}=\nu_i P_{ij}$$

因为ν_i是过程离开状态i的速率而P_{ij}是它转移到j的概率，所以q_{ij}是过程从状态i转移到状态j的速率；事实上我们就称q_{ij}是从i到j的转移速率。

以$P_{ij}(t)$记马尔可夫链现在处于状态i，再经过一段时间t后

处于状态 j 的概率，即

$$P_{ij}(t) = P\{X(t+s) = j \mid X(s) = i\}$$

5.3 生灭过程

具有状态 0，1，…的连续时间马尔可夫链称为生灭过程，若 $|i-j|>1$ 时 $q_{ij}=0$。于是一个生灭过程是一个连续时间马尔可夫链，具有状态 0，1…，它从状态 i 只能转移到状态 $i-1$ 或 $i+1$。过程的状态通常看作为某个群体的总量，当状态增长 1 时，我们就说生了一个；而当它减少 1 时，我们就说死了一个。设 λ_i 与 μ_i 为

$$\lambda_i = q_{i,i+1}$$

$$\mu_i = q_{i,i-1}$$

值$\{\lambda_i, i\geqslant 0\}$与$\{\mu_i, i\geqslant 1\}$分别称为生长率与死亡率。因为$\sum_{j\neq i} q_{ij} = \nu_i$，可见

$$\nu_i = \lambda_i + \mu_i$$

$$P_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i} = 1 - P_{i,i-1}$$

因此，我们可以这样设想生灭过程，每当系统中有 i 个人时，直到下一次出生的时间服从参数为 λ_i 的指数分布，且独立于直到下一次死亡的时间，它服从参数为 μ_i 的指数分布。

例 5.3 (a)　两个生灭过程。

(i) $M/M/s$ 排队系统。假设顾客按照参数为 λ 的泊松过程来到一个有 s 个服务员的服务站，即相继来到之间的时间是均值为 $1/\lambda$ 的独立指数随机变量，每个顾客一来到，如果有服务员空闲，则直接进入服务，否则此顾客要加入排队行列(即他在队列中等待)。当一个服务员结束对一位顾客的服务时，顾客便离开服务系统，排队中的下一个顾客(若有顾客在等待)进入服务。假定相继的服务时间是独立的指数随机变量，均值为 $1/\mu$。如果我们以 $X(t)$ 记时刻 t 系统中的人数，则$\{X(t), t\geqslant 0\}$是生灭过程，

$$\mu_n = \begin{cases} n\mu, & 1 \leqslant n \leqslant s \\ s\mu, & n > s \end{cases}$$

$$\lambda_n = \lambda, \qquad n \geqslant 0$$

(ii) 有迁入的线性增长模型

$$\mu_n = n\mu, \qquad n \geqslant 1$$

$$\lambda_n = n\lambda + \theta, \qquad n \geqslant 0$$

的模型称为有迁入的线性增长模型。这种过程自然地产生于生物繁殖与群体增长的研究中。假定群体中的每个个体以指数率 λ 出生；此外，群体由于从外界迁入的因素又以指数率 θ 增加，因此在系统中有 n 人时，整个出生率是 $n\lambda+\theta$。假定此群体的各个成员以指数率 μ 死亡，从而 $\mu_n=n\mu$。

若对一切 n，$\mu_n=0$（即若死亡是不可能的），则生灭过程称为纯生过程。最简单的纯生过程的例子是泊松过程，它具有常值出生率 $\lambda_n=\lambda$，$n\geqslant 0$。

第二个纯生过程的例子是这样的，在一个群体中各个成员独立地活动且以指数率 λ 生育。若假设没有任何成员死亡，以 $X(t)$ 记时刻 t 群体的总量，则 $\{X(t), t\geqslant 0\}$ 是一个纯生过程，其

$$\lambda_n = n\lambda, \qquad n \geqslant 0$$

此纯生过程被称为尤尔过程。

考虑一尤尔过程，在时刻 0 从一个个体开始，且以 T_i $(i\geqslant 1)$ 记第 $i-1$ 个与第 i 个出生之间的时间。即 T_i 是群体总量从 i 变到 $i+1$ 所花的时间。从尤尔过程的定义容易得到 T_i $(i\geqslant 1)$ 是独立的，且 T_i 是具有参数 $i\lambda$ 的指数变量。现在

$$P\{T_1 \leqslant t\} = 1 - e^{-\lambda t}$$

$$\begin{aligned}
P\{T_1 + T_2 \leqslant t\} &= \int_0^t P\{T_1 + T_2 \leqslant t \mid T_1 = x\} \lambda e^{-\lambda x} dx \\
&= \int_0^t (1 - e^{-2\lambda(t-x)}) \lambda e^{-\lambda x} dx \\
&= (1 - e^{-\lambda t})^2
\end{aligned}$$

$$\begin{aligned}
P\{T_1 + T_2 + T_3 \leqslant t\} &= \int_0^t P\{T_1 + T_2 + T_3 \leqslant t \mid T_1 + T_2 \\
&\qquad = x\} dF_{T_1+T_2}(x) \\
&= \int_0^t (1 - e^{-3\lambda(t-x)}) 2\lambda e^{-\lambda x}(1 - e^{-\lambda x}) dx
\end{aligned}$$

$$= (1 - e^{-\lambda t})^3$$

一般地可用归纳法证明

$$P\{T_1 + \cdots + T_j \leqslant t\} = (1 - e^{-\lambda t})^j$$

因此，由 $P\{T_1+\cdots+T_j\leqslant t\}=P\{X(t)\geqslant j+1|X(0)=1\}$ 可见对于一个尤尔过程，

$$P_{1j}(t) = (1 - e^{-\lambda t})^{j-1} - (1 - e^{-\lambda t})^j$$
$$= e^{-\lambda t}(1 - e^{-\lambda t})^{j-1}, \qquad j \geqslant 1$$

从上可见，从一个个体开始，在时刻 t 群体的总量有几何分布，其均值为 $e^{\lambda t}$。因此如果群体从 i 个个体开始，在时刻 t 其总量是 i 个独立同几何分布随机变量之和，有负二项分布，也即对尤尔过程

$$P_{ij}(t) = \binom{j-1}{i-1} e^{-\lambda t i}(1 - e^{-\lambda t})^{j-i}, \qquad j \geqslant i \geqslant 1$$

关于从一个个体开始的尤尔过程的另一个有趣的结果涉及时刻 t 的群体总量给定时出生时刻的条件分布。因为第 i 个出生在时刻 $S_i\equiv T_1+\cdots+T_i$ 发生，所以我们计算已给 $X(t)=n+1$ 时 S_1，$\cdots$，S_n 的条件联合分布。直观地推导，并将密度当作概率处理可得，对 $0\leqslant s_1\leqslant s_2\leqslant\cdots\leqslant s_n\leqslant t$

$$P\{S_1 = s_1, S_2 = s_2, \cdots, S_n = s_n | X(t) = n + 1\}$$
$$= \frac{P\{T_1 = s_1, T_2 = s_2 - s_1, \cdots, T_n = s_n - s_{n-1}, T_{n+1} > t - s_n\}}{P\{X(t) = n + 1\}}$$
$$= \frac{\lambda e^{-\lambda s_1} 2\lambda e^{-2\lambda(s_2 - s_1)} \cdots n\lambda e^{-n\lambda(s_n - s_{n-1})} e^{-(n+1)\lambda(t - s_n)}}{P\{X(t) = n + 1\}}$$
$$= Ce^{-\lambda(t-s_1)} e^{-\lambda(t-s_2)} \cdots e^{-\lambda(t-s_n)}$$

其中 C 是某个不依赖于 s_1，$\cdots s_n$ 的常数。因此我们看到，给定 $X(t)=n+1$ 时 S_1，$\cdots$，S_n 的条件密度为

(5.3.1)

$$f(s_1, \cdots, s_n | n+1) = n! \prod_{i=1}^{n} f(s_i), \qquad 0 \leqslant s_1 \leqslant \cdots \leqslant s_n \leqslant t$$

其中 f 是密度函数

(5.3.2)

$$f(x)=\begin{cases}\dfrac{\lambda e^{-\lambda(t-x)}}{1-e^{-\lambda t}}, & 0\leqslant x\leqslant t\\ 0, & \text{其它}\end{cases}$$

但是因为(5.3.1)是 n 个密度为 f 的随机变量的一个子样的顺序统计量的联合密度函数(参阅第二章 2.3 节)。于是我们证得

命题 5.3.1

考虑一个尤尔过程，其 $X(0)=1$，则给定 $X(t)=n+1$ 时，出生时刻 S_1，S_2，…，S_n 的分布如同取自密度为(5.3.2) 的母体的容量为 n 的子样的顺序统计量的分布。

命题 5.3.1 可用来以同样的方法对尤尔过程建立与泊松过程相应的结果。

例 5.3(b) 考虑一尤尔过程，其 $X(0)=1$。让我们计算在时刻 t 群体诸成员的年龄之和的均值。时刻 t 诸年龄之和，记为 $A(t)$，可表示为

$$A(t)=a_0+t+\sum_{i=1}^{x(t)-1}(t-S_i)$$

其中 a_0 是初始个体在 $t=0$ 时的年龄。为计算 $E[A(t)]$，对 $X(t)$ 取条件

$$\begin{aligned}E[A(t)|X(t)=n+1]&=a_0+t+E[\sum_{i=1}^{n}(t-S_i)|X(t)=n+1]\\&=a_0+t+n\int_0^t(t-x)\frac{\lambda e^{-\lambda(t-x)}}{1-e^{-\lambda t}}dx\end{aligned}$$

或

$$E[A(t)|X(t)]=a_0+t+(X(t)-1)\frac{1-e^{-\lambda t}-\lambda te^{-\lambda t}}{\lambda(1-e^{-\lambda t})}$$

取期望且由 $X(t)$ 有均值 $e^{\lambda t}$ 得

$$\begin{aligned}E[A(t)]&=a_0+t+\frac{e^{\lambda t}-1-\lambda t}{\lambda}\\&=a_0+\frac{e^{\lambda t}-1}{\lambda}\end{aligned}$$

其它与 $A(t)$ 有关的量，例如它的母函数可按同样的方法计算。

上面 $E[A(t)]$ 的公式可用下面的恒等式加以验证，此式的证明留作一个练习：

(5.3.3) $$A(t)=a_0+\int_0^t X(s)ds$$

取期望得

$$E[A(t)]=a_0+E\left[\int_0^t X(s)ds\right]$$

$$=a_0+\int_0^t E[X(s)]ds \qquad (\text{因为 } X(s)\geqslant 0)$$

$$=a_0+\int_0^t e^{\lambda s}ds$$

$$=a_0+\frac{e^{\lambda t}-1}{\lambda}$$

下面的例子提供了纯生过程的另一种解释。

例 5.3(c)　一个简单的传染模型。考虑有 m 个个体的群体，在时刻 0 由一个已感染的个体与 $m-1$ 个未受到感染但能被感染的个体组成。个体一旦受到感染将永远地处于此状态。假设在任意的长为 h 的时间区间内任意一个已感染的人将以概率 $\alpha h+o(h)$ 引起任一指定的未被感染者成为已感染者。若我们以 $X(t)$ 记时刻 t 群体中已受感染的个体数，则 $\{X(t),t\geqslant 0\}$ 是一纯生过程，其

$$\lambda_n=\begin{cases}(m-n)n\alpha, & n=1,\cdots,m-1\\ 0, & \text{其它}\end{cases}$$

这是因为当有 n 个已受感染的个体时则 $m-n$ 个未受感染者的每一个将以速率 $n\alpha$ 变成已感染者。

以 T 记直至整个群体被感染的时间，则 T 能表示为

$$T=\sum_{i=1}^{m-1}T_i$$

其中 T_i 是从 i 个已感染者到 $i+1$ 个已感染者的时间。因为 T_i 是独立指数随机变量，其参数分别为 $\lambda_i=(m-i)i\alpha$，$i=1$，$\cdots$，$m-1$，可见

$$E[T]=\frac{1}{\alpha}\sum_{i=1}^{m-1}\frac{1}{i(m-i)}$$

及

$$\mathrm{Var}(T)=\frac{1}{\alpha^2}\sum_{i=1}^{m-1}\left(\frac{1}{i(m-i)}\right)^2$$

对规模合理的群体，$E[T]$ 渐近地为

$$E[T]=\frac{1}{m\alpha}\sum_{i=1}^{m-1}\left(\frac{1}{m-i}+\frac{1}{i}\right)$$

$$\approx \frac{1}{m\alpha}\int_1^{m-1}\left(\frac{1}{m-t}+\frac{1}{t}\right)dt=\frac{2\log(m-1)}{m\alpha}$$

5.4 柯尔莫哥洛夫微分方程

记得

$$P_{ij}(t)=P\{X(t+s)=j|X(s)=i\}$$

代表过程目前处于状态 i 在时间 t 之后将处于状态 j 的概率。

利用马尔可夫性，我们将导出两组 $P_{ij}(t)$ 的微分方程，它们有时可求得显式解。但在此之前需要下面的引理。

引理 5.4.1

(i) $\lim\limits_{t\to 0}\frac{1-P_{ij}(t)}{t}=\nu_i$

(ii) $\lim\limits_{t\to 0}\frac{P_{ij}(t)}{t}=q_{ij}, \quad i\neq j$

引理 5.4.2

对一切 s，t，

$$P_{ij}(t+s)=\sum_{k=0}^{\infty}P_{ik}(t)P_{kj}(s)$$

引理 5.4.1 从如下事实(它们必须加以证明)可得，在时间 t 内有两次或更多次转移的概率是 $o(t)$；而引理 5.4.2，它是离散时间马尔可夫链的切普曼——柯尔莫哥洛夫方程的连续时间的翻版，直接从马尔可夫性来推得。证明的细节留作练习。

从引理 5.4.2 我们得到

$$P_{ij}(t+h)=\sum_{k=0}^{\infty}P_{ik}(h)P_{kj}(t)$$

或等价地，

$$P_{ij}(t+h)-P_{ij}(t)=\sum_{k\neq i}P_{ik}(h)P_{kj}(t)-[1-P_{ii}(h)]P_{ij}(t)$$

除以 h 而后令 $h\to 0$ 取极限，应用引理 5.4.1 得

(5.4.1)

$$\lim_{h\to 0}\frac{P_{ij}(t+h)-P_{ij}(t)}{h}=\lim_{h\to 0}\sum_{k\neq i}\frac{P_{ik}(h)}{h}P_{kj}(t)-\nu_i P_{ij}(t)$$

假定在（5.4.1）的右边可交换极限与求和，再用引理 5.4.1，于是得到下面结论。

定理 5.4.3（柯尔莫哥洛夫向后方程）对一切 i，j 及 $t\geqslant 0$，

$$P'_{ij}(t)=\sum_{k\neq i}q_{ik}P_{kj}(t)-\nu_i P_{ij}(t)$$

证明 为完成证明必须论证（5.4.1）右边极限与求和可交换次序。现在，对于任意固定的 N，

$$\liminf_{h\to 0}\sum_{k\neq i}\frac{P_{ik}(h)}{h}P_{kj}(t)\geqslant\liminf_{h\to 0}\sum_{\substack{k\neq i\\k<N}}\frac{P_{ik}(h)}{h}P_{kj}(t)$$

$$=\sum_{\substack{k\neq i\\k<N}}q_{ik}P_{kj}(t)$$

因为上式对一切 N 成立，可见

(5.4.2)

$$\lim_{h\to 0}\inf\sum_{k\neq i}\frac{P_{ik}(h)}{h}P_{kj}(t)\geqslant\sum_{k\neq i}q_{ik}P_{kj}(t)$$

为了倒转不等式，注意对于 $N>i$，由于 $P_{kj}(t)\leqslant 1$，所以

$$\begin{aligned}&\limsup_{h\to 0}\sum_{k\neq i}\frac{P_{ik}(h)}{h}P_{kj}(t)\\&\leqslant\limsup_{h\to 0}\left[\sum_{\substack{k\neq i\\k<N}}\frac{P_{ik}(h)}{h}P_{kj}(t)+\sum_{k\geqslant N}\frac{P_{ik}(h)}{h}\right]\\&\leqslant\limsup_{h\to 0}\left[\sum_{\substack{k\neq i\\k<N}}\frac{P_{ik}(h)}{h}P_{kj}(t)+\frac{1-P_{ii}(h)}{h}-\sum_{\substack{k\neq i\\k<N}}\frac{P_{ik}(h)}{h}\right]\\&=\sum_{\substack{k\neq i\\k<N}}q_{ik}P_{kj}(t)+\nu_i-\sum_{\substack{k\neq i\\k<N}}q_{ik}\end{aligned}$$

其中最后的等式由引理 5.4.1 而得。因上列不等式对一切 $N>i$ 成立，令 $N\to\infty$ 且用 $\sum_{k\neq i}q_{ik}=\nu_i$，我们即得

$$\limsup_{h\to 0}\sum_{k\neq i}\frac{P_{ik}(h)}{h}P_{kj}(t)\leqslant\sum_{k\neq i}q_{ik}P_{kj}(t)$$

上式连同(5.4.2)证明了

$$\lim_{h\to 0}\sum_{k\neq i}\frac{P_{ik}(h)}{h}P_{kj}(t)=\sum_{k\neq i}q_{ik}P_{kj}(t)$$

定理 5.4.3 得证。

定理 5.4.3 中 $P_{ij}(t)$ 满足的微分方程组以柯尔莫哥洛夫向后方程著称。称它们为向后方程，是因为在计算时刻 $t+h$ 的状态的概率分布时我们对退后到时刻 h 的状态取条件，即我们从

$$\begin{aligned}P_{ij}(t+h)=&\sum_{k=0}^{\infty}P\{X(t+h)=j\,|\,X(0)=i,\\&X(h)=k\}P\{X(h)=k\,|\,X(0)=i\}\\=&\sum_{k=0}^{\infty}P_{kj}(t)P_{ik}(h)\end{aligned}$$

开始计算。

对时刻 t 的状态取条件，我们可以导出另一组方程，称柯尔莫哥洛夫向前方程。可得

$$P_{ij}(t+h)=\sum_{k=0}^{\infty}P_{ik}(t)P_{kj}(h)$$

或

$$\begin{aligned}P_{ij}(t+h)-P_{ij}(t)&=\sum_{k=0}^{\infty}P_{ik}(t)P_{kj}(h)-P_{ij}(t)\\&=\sum_{k\neq j}P_{ik}(t)P_{kj}(h)-[1-P_{jj}(h)]P_{ij}(t)\end{aligned}$$

所以，

$$\begin{aligned}&\lim_{h\to 0}\frac{P_{ij}(t+h)-P_{ij}(t)}{h}\\&=\lim_{h\to 0}\left\{\sum_{k\neq j}P_{ik}(t)\frac{P_{kj}(h)}{h}-\frac{1-P_{jj}(h)}{h}P_{ij}(t)\right\}\end{aligned}$$

假定我们能交换极限与求和，由引理 5.4.1 便得到

$$P'_{ij}(t)=\sum_{k\neq j}q_{kj}P_{ik}(t)-\nu_jP_{ij}(t)$$

令人遗憾的是并非一定能论证极限与求和可交换，所以上式并非总是成立的。然而，在大多数模型中——包括全部生灭过程与全部有限状态的模型，它们确实是成立的。于是有

定理 5.4.4（柯尔莫哥洛夫向前方程）。在适当的正则条件下，

$$P_{ij}(t)=\sum_{k\neq j}q_{kj}P_{ik}(t)-\nu_j P_{ij}(t)$$

例 5.4（a）　两状态链。考虑两个状态的连续时间马尔可夫链，在转移到状态 1 之前链在状态 0 停留的时间是参数为 λ 的指数变量，而在回到状态 0 之前它停留在状态 1 的时间是参数为 μ 的指数变量，向前方程为

$$\begin{aligned}P'_{00}(t)&=\mu P_{01}(t)-\lambda P_{00}(t)\\&=-(\lambda+\mu)P_{00}(t)+\mu\end{aligned}$$

其中最后的等式来自 $P_{01}(t)=1-P_{00}(t)$。因此

$$e^{(\lambda+\mu)t}[P'_{00}(t)+(\lambda+\mu)P_{00}(t)]=\mu e^{(\lambda+\mu)t}$$

或

$$\frac{d}{dt}[e^{(\lambda+\mu)t}P_{00}(t)]=\mu e^{(\lambda+\mu)t}$$

于是

$$e^{(\lambda+\mu)t}P_{00}(t)=\frac{\mu}{\lambda+\mu}e^{(\lambda+\mu)t}+c$$

由于 $P_{00}(0)=1$，可见 $c=\lambda\ (\lambda+\mu)$，于是

$$P_{00}(t)=\frac{\mu}{\lambda+\mu}+\frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t}$$

类似地（或由对称性），

$$P_{11}(t)=\frac{\lambda}{\lambda+\mu}+\frac{\mu}{\lambda+\mu}e^{-(\lambda+\mu)t}$$

例 5.4（b）　生灭过程的柯尔莫哥洛夫向前方程是

$$P'_{i0}(t)=\mu_1P_{i1}(t)-\lambda_0P_{i0}(t)$$

$$P'_{ij}(t)=\lambda_{j-1}P_{i,j-1}(t)+\mu_{j+1}P_{i,j+1}(t)-(\lambda_j+\mu_j)P_{ij}(t),\qquad j\neq 0$$

例 5.4（c）　对纯生过程，向前方程归结为

(5.4.3)

$$P'_{ii}(t)=-\lambda_iP_{ii}(t)$$

$$P'_{ij}(t)=\lambda_{j-1}P_{i,j-1}(t)-\lambda_jP_{ij}(t),\qquad j>i$$

积分(5.4.3)的第一个方程并用 $P_{ii}(0)=1$ 得

$$P_{ii}(t)=e^{-\lambda_i t}$$

上式当然是正确的，因为 $P_{ii}(t)$ 是从状态 i 转移出去的时间超过 t 的概率。其它的量 $P_{ij}(t)$，$j>i$，能从(5.4.3)如下递推而得：对 $j>i$，从(5.4.3)有

$$e^{\lambda_j t}\lambda_{j-1}P_{i,j-1}(t)=e^{\lambda_j t}[P'_{ij}(t)+\lambda_jP_{ij}(t)]$$

$$=\frac{d}{dt}[e^{\lambda_j t}P_{ij}(t)]$$

积分并用 $P_{ij}(0)=0$ 得

$$P_{ij}(t)=\lambda_{j-1}e^{-\lambda_j t}\int_0^t e^{\lambda_j s}P_{i,j-1}(s)ds, \qquad j>i$$

在尤尔过程的特殊情形，其中 $\lambda_j=j\lambda$，可用上式验证 5.3 节的结果，即有

$$P_{ij}(t)=\binom{j-1}{i-1}e^{-\lambda t i}(1-e^{-\lambda t})^{j-i}, \qquad j\geqslant i\geqslant 1$$

注记　若定义 $q_{ii}\equiv-\nu_i$，则向后方程可写为

$$P'_{ij}(t)=\sum_{k=0}^{\infty}q_{ik}P_{kj}(t)$$

而向前方程为

$$P'_{ij}(t)=\sum_{k=0}^{\infty}q_{kj}P_{ik}(t)$$

这些方程用矩阵记号写有特别好的形式。若我们定义$\underline{P}(t)$，$\underline{Q}$ 及 $\underline{P}'(t)$为矩阵，其(i,j)位置上的元素分别是 $P_{ij}(t)$，q_{ij}及 $P'_{ij}(t)$，则向后方程能写为

$$\underline{P}'(t)=\underline{Q}\,\underline{P}(t)$$

而向前方程为

$$\underline{P}'(t)=\underline{P}(t)\underline{Q}$$

如果$\underline{P}(t)$只是 t 的函数而 Q 是一个常数，则上式的解是

$$P(t)=e^{Qt}=\sum_{j=0}^{\infty}\frac{(Qt)^i}{i!}$$

因此认为

$$\underline{P}(t)=\sum_{i=0}^{\infty}\frac{t^i\,\underline{Q}^i}{i!} \tag{5.4.4}$$

是$\underline{P}(t)$的可能解似乎是合理的，其中$\underline{Q}^0=I$（单位矩阵）。事实上能够证明，当 ν_i 有界时(5.4.4)是成立的。因此，当状态空间有限时，(5.4.4)成立且事实上这可能是一种近似计算$\underline{P}(t)$的方便的方法(另外的逼近方法在 5.8 节中出现)。

5.5 极限概率

因为连续时间马尔可夫链是一个半马尔可夫过程，其

$$F_{ij}(t)=1-e^{-\nu_i t}$$

从第四章 4.8 节的结果得，如果转移概率为 P_{ij}的离散时间马尔可夫链是不可约正常返的，则极限概率 $P_j=\lim_{t\to\infty} P_{ij}(t)$ 为

$$P_j=\frac{\pi_j/\nu_j}{\sum_{i=0}^{\infty}\pi_i/\nu_i} \tag{5.5.1}$$

其中 π_j 是

$$\pi_j=\sum_{i=0}^{\infty}\pi_i P_{ij} \tag{5.5.2}$$

$$\sum_{i=0}^{\infty}\pi_i=1$$

的唯一非负解。从(5.5.1)与(5.5.2)可见 P_j 是

$$\nu_j P_j=\sum_{i=0}^{\infty}\nu_i P_i P_{ij}$$

$$\sum_{j=0}^{\infty}P_j=1$$

或等价地，用 $q_{ij}=\nu_i P_{ij}$，是

$$\nu_j P_j=\sum_{i=0}^{\infty}P_i P_{ij} \tag{5.5.3}$$

$$\sum_{j=0}^{\infty}P_j=1$$

的唯一非负解。

注记

(1) 从第四章 4.8 节中给出的关于半马尔可夫过程的结果得出，P_j 也等于长时间之后过程处于状态 j 的时间的比率。

(2) 如果初始状态按照极限概率 $\{P_j\}$ 选取，则所得过程将是

平稳的，即对一切 t $\quad \sum_{i=0}^{\infty} P_i P_{ij}(t) = P_j$

上式的证明如下：

$$\begin{aligned}\sum_{i=0}^{\infty} P_{ij}(t) P_i &= \sum_{i=0}^{\infty} P_{ij}(t) \lim_{s \to \infty} P_{ki}(s) \\ &= \lim_{s \to \infty} \sum_{i=0}^{\infty} P_{ij}(t) P_{ki}(s) \\ &= \lim_{s \to \infty} P_{kj}(t+s) \\ &= P_j\end{aligned}$$

容易论证上式的求和与极限可交换，将此留作一个练习。

(3)得到方程(5.5.3)的另一途径是用向前方程

$$P_{ij}(t) = \sum_{k \neq j} q_{kj} P_{ik}(t) - \nu_j P_{ij}(t)$$

若我们假定极限概率 $P_j = \lim_{t \to \infty} P_{ij}(t)$ 存在，则 $t \to \infty$ 时，$P_{ij}'(t)$ 必收敛于 0。(为什么?)因此，假定上式中可以交换极限与求和号，令 $t \to \infty$ 即得

$$0 = \sum_{k \neq j} P_k q_{kj} - \nu_j P_j$$

值得注意的是上述方法是下列直观论证的较正式的叙述。为了得到 P_j（$t=\infty$时处于状态 j 的概率）的方程，对 h 个单位时间之前的状态取条件：

$$\begin{aligned}P_j &= \sum_{i=0}^{\infty} P_{ij}(h) P_i \\ &= \sum_{i \neq j} (q_{ij} h + o(h)) P_i + (1 - \nu_j h + o(h)) P_j\end{aligned}$$

或

$$0 = \sum_{i \neq j} P_i q_{ij} - \nu_j P_j + \frac{o(h)}{h}$$

令 $h \to 0$ 即得结果。

(4) 方程(5.5.3)有一种很好的解释如下：在任何区间$(0,t)$中，转移到状态 j 的次数与从状态 j 转移出来的次数相差不超过 1

(为什么?)因此,长时间之后转移到状态 j 发生的速率必等于从状态 j 转移出去发生的速率。既然过程处于状态 j 时它以速率ν_j离开,又因 P_j 是过程处于状态 j 的时间的比率,于是得

$$\nu_j P_j = \text{过程离开状态 } j \text{ 的速率}$$

类似地,当过程处于状态 i 时它以速率 q_{ij} 离开而转到 j,又因 P_i 是处于状态 i 的时间的比率,可见从 i 到 j 的转移发生的速率等于 $q_{ij}P_i$。因此,

$$\sum_{i=0}^{\infty} P_i q_{ij} = \text{过程进入状态 } j \text{ 的速率}$$

所以,(5.5.3)正是说过程进入与离开状态 j 的速率相等。因为它使这些速率平衡(即相等),所以方程(5.5.3)有时称为平衡方程。

(5)当连续时间马尔可夫链不可约且对一切 j 有 $P_j>0$ 时,我们说链是遍历的。

现在让我们对生灭过程确定其极限概率。从方程(5.5.3),或等价地,使过程离开一个状态与进入该状态的速率相等,得到

状态	过程离开的速率		过程进入的速率
0	$\lambda_0 P_0$	=	$\mu_1 P_1$
n, $n>0$	$(\lambda_n+\mu_n)P_n$	=	$\mu_{n+1}P_{n+1}+\lambda_{n-1}P_{n-1}$

改写这些方程为

$$\lambda_0 P_0 = \mu_1 P_1$$

$$\lambda_n P_n = \mu_{n+1}P_{n+1} + (\lambda_{n-1}P_{n-1} - \mu_n P_n), \qquad n \geqslant 1$$

或等价地

$$\lambda_0 P_0 = \mu_1 P_1$$

$$\lambda_1 P_1 = \mu_2 P_2 + (\lambda_0 P_0 - \mu_1 P_1) = \mu_2 P_2$$

$$\lambda_2 P_2 = \mu_3 P_3 + (\lambda_1 P_1 - \mu_2 P_2) = \mu_3 P_3$$

$$\lambda_n P_n = \mu_{n+1}P_{n+1} + (\lambda_{n-1}P_{n-1} - \mu_n P_n) = \mu_{n+1}P_{n+1}$$

用 P_0 解得

$$P_1=\frac{\mu_0}{\lambda_1}P_0$$

$$P_2=\frac{\lambda_1}{\mu_2}P_1=\frac{\lambda_1\lambda_0}{\mu_2\mu_1}P_0$$

$$P_3=\frac{\lambda_2}{\mu_3}P_2=\frac{\lambda_2\lambda_1\lambda_0}{\mu_3\mu_2\mu_1}P_0$$

$$P_n=\frac{\lambda_{n-1}}{\mu_n}P_{n-1}=\frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_1\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_2\mu_1}P_0$$

利用$\sum_{n=0}^{\infty}P_n=1$得

$$1=P_0+P_0\sum_{n=1}^{\infty}\frac{\lambda_{n-1}\cdots\lambda_1\lambda_0}{\mu_n\cdots\mu_2\mu_1}$$

或

$$P_0=\left[1+\sum_{n=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}\right]^{-1}$$

因此

$$P_n=\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n\left(1+\sum_{n=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}\right)},\qquad n\geqslant 1 \tag{5.5.4}$$

上式也给我们指明了极限概率存在所需条件，即

$$\sum_{n=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}<\infty$$

例 5.5 (a)　***M/M/*1 排队系统**。在 $M/M/1$ 排队系统中，$\lambda_n=\lambda$，$\mu_n=\mu$，于是若 $\lambda/\mu<1$ 从(5.5.4)得

$$P_n=\frac{(\lambda/\mu)^n}{1+\sum_{n=1}^{\infty}\left(\frac{\lambda}{\mu}\right)^n}=\left(\frac{\lambda}{\mu}\right)^n\left(1-\frac{\lambda}{\mu}\right),\qquad n\geqslant 0$$

要极限概率存在，λ 必须小于 μ 是直观的。顾客们按速率 λ 到来且以速率 μ 受到服务，因而当 $\lambda>\mu$ 时他们到来的速率高于他们能受到服务的速率，排队长度将趋于无穷。$\lambda=\mu$ 的情况很象第四章 4.3 节的对称随机游动，它是零常返的，从而没有极限概率。

例 5.5 (b)　考虑一个有 M 部机器与一个修理工的车间，且假设一部

机器在损坏前运转的时间服从参数为 λ 的指数分布，而修理工修好一部损坏了的机器的时间服从参数为 μ 的指数分布。每当有 n 部机器坏了就说状态为 n，则此系统为一个生灭过程的模型，其参数

$$\mu_n=\mu, \qquad n\geqslant 1$$

$$\lambda_n=\begin{cases}(M-n)\ \lambda, & n\leqslant M\\ 0, & n>M\end{cases}$$

从方程(5.5.4)可得 n 部机器不在使用的极限概率 P_n 为

$$P_0=\frac{1}{1+\sum_{n=1}^{M}\left(\frac{\lambda}{\mu}\right)^n\frac{M!}{(M-n)!}}$$

$$P_n=\frac{\frac{M!}{(M-n)!}\left(\frac{\lambda}{\mu}\right)^n}{1+\sum_{n=1}^{M}\left(\frac{\lambda}{\mu}\right)^n\frac{M!}{(M-n)!}}, \qquad n=0,\ \cdots,\ M$$

因此，不在使用的机器的平均台数为

$$\sum_{n=0}^{M}nP_n=\frac{\sum_{n=0}^{M}n\frac{M!}{(M-n)!}\left(\frac{\lambda}{\mu}\right)^n}{1+\sum_{n=1}^{M}\left(\frac{\lambda}{\mu}\right)^n\frac{M!}{(M-n)!}}$$

假设我们想知道长时间之后一部指定的机器在工作的时间的比，为此等价地我们计算它在工作的极限概率：

$$\begin{aligned}P\{\text{该机器在工作}\}&=\sum_{n=0}^{M}P\{\text{该机器在工作}\mid n\text{ 部未工作}\}P_n\\&=\sum_{n=0}^{M}\frac{M-n}{M}P_n\end{aligned}$$

5.6 时间可逆性

考虑一个遍历的连续时间马尔可夫链，且假定已进行了无限长时间；例如，假定它从时间 $t=-\infty$ 开始。这样一个过程将是平稳的，我们说它处于稳态之中。（产生平稳过程的另一方法是假设在时刻 $t=0$ 的初始状态按极限概率选取）在时刻 t 开始让我们把时间方向倒过来跟踪这过程。为了确定这逆向过程的概率结构，首

先我们注意到在某时刻(譬如说 t)处于状态 i 的条件下,处于此状态的时间已超过 s 的概率正是 $e^{-\nu_i s}$。这是因为

$$
\begin{aligned}
&P\{[t-s,t]\text{内过程全处于状态 } i \mid X(t)=i\} \\
&\quad = P\{[t-s,t]\text{内过程全处于状态 } i\}/P\{X(t)=i\} \\
&\quad = \frac{P\{X(t-s)=i\}e^{-\nu_i s}}{P\{X(t)=i\}} \\
&\quad = e^{-\nu_i s}
\end{aligned}
$$

其中 $P\{X(t-s)=i\}=P\{X(t)=i\}=P_i$。

换言之,逆时间退回去,过程处于状态 i 的时间也服从指数分布,参数为 ν_i。此外,如同在第四章 4.7 节中证明的,逆向过程所到达的状态序列构成一个离散时间马尔可夫链,其转移概率 Q_{ij} 为

$$Q_{ij}=\frac{\pi_j P_{ji}}{\pi_i}$$

因此从上式可见逆向过程是一个连续时间马尔可夫链,离开每一个状态的转移速率与正向过程相同,一步转移概率为 Q_{ij}。所以,连续时间马尔可夫链在时间逆向过程与原过程具有同样的概率结构的意义下是时间可逆的,如果嵌入链是时间可逆的—— 即如果对一切 i,j

$$\pi_i P_{ij}=\pi_j P_{ji}$$

现在利用

$$P_i=\frac{\pi_i/\nu_i}{\sum_{j=0}^{\infty}\frac{\pi_j}{\nu_j}}$$

可见上面的条件等价于对一切 $i\neq j$

$$P_i\nu_i P_{ij}=P_j\nu_j P_{ji}$$

或等价地,对一切 $i\neq j$

$$P_i q_{ij}=P_j q_{ji} \tag{5.6.1}$$

由于 P_i 是处于状态 i 的时间的比率,且处于状态 i 的过程以速率 q_{ij} 转移到 j,所以时间可逆条件就是过程直接从状态 i 到状态 j 的速率等于它直接从 j 到 i 的速率。应当注意到这正是遍历的离散

马尔可夫链时间可逆所需的同一条件(见第四章 4.7 节)。

应用上述时间可逆的条件得出有关生灭过程的下列命题。

命题 5.6.1

遍历生灭过程在稳态下是时间可逆的。

证明 为证此命题,我们必须证明生灭过程从状态 i 到状态 $i+1$ 的速率等于它从 $i+1$ 到 i 的速率。既然在任意长为 t 的时间内从 i 转移到 $i+1$ 的次数与从 $i+1$ 转移到 i 的次数相差不过 1(因为从 i 到 $i+1$ 的两次转移之间过程必须返回到 i,且只能通过 $i+1$ 回到 i,反之亦然)。从而由于 $t\to\infty$ 时这种转移的次数趋于无穷,得从 i 到 $i+1$ 的转移的速率等于从 $i+1$ 到 i 的速率。

命题 5.6.1 可用于证明 $M/M/s$ 系统的输出过程是泊松过程,把它叙述为一个系。

系 5.6.2

考虑一个 $M/M/s$ 系统,其中顾客依照速率为 λ 的泊松过程来到,且受到 s 个服务员中任一个的服务——每一个服务时间服从参数为 μ 的指数分布。若 $\lambda<s\mu$,则顾客离开的输出过程在稳态下是速率为 λ 的泊松过程。

证明 以 $X(t)$ 记时刻 t 系统中的顾客数。因为 $M/M/s$ 过程是一生灭过程,从命题 5.6.1 得 $\{X(t),t\geqslant 0\}$ 是时间可逆的。依正向时间向前,$X(t)$ 增加 1 的时刻组成一泊松过程,因为它们正是顾客来到的时刻,因此由时间可逆性,按逆向时间 $X(t)$ 增加 1 的时刻也组成一泊松过程。但是它们正好是原过程顾客离开的时刻(见图 5.6.1)。因此离去的时刻构成一速率为 λ 的泊松过程。

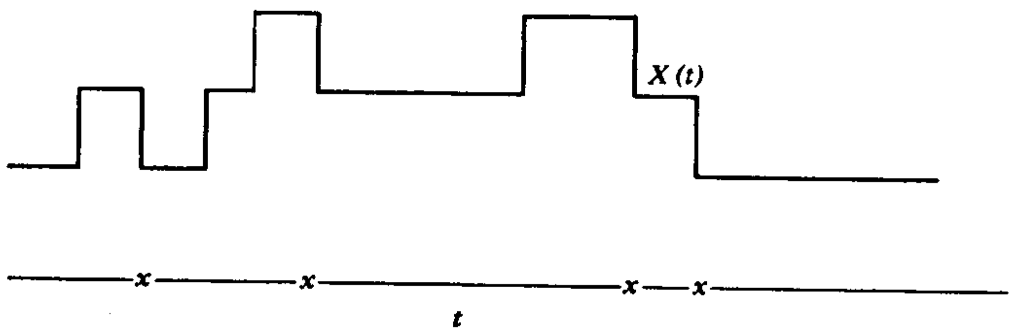


图 5.6.1 x 为按逆向时间 $X(t)$ 增长的时刻;
x 也等于按正向时间 $X(t)$ 减少的时刻。

考虑一个连续时间马尔可夫链，其状态空间是 S。我们说马尔可夫链被截于集 $A\subset S$，若对一切 $i\in A$，$j\notin A$ 改变 $q_{ij}=0$，所有其它的 q_{ij} 保留不变。于是从状态类 A 中转移出来是不允许的。一个有用的结果是被截得的时间可逆链仍然是时间可逆的。

命题 5.6.3

具有极限概率 P_j ($j\in S$) 的时间可逆链，若截于集 $A\subset S$，被截所得仍不可约，则也是时间可逆的且有极限概率。

(5.6.2)
$$P_j^A=P_j\Big/\sum_{j\in A}P_j,\qquad j\in A$$

证明 我们必须证明，对 $i\in A$，$j\in A$

$$P_i^A q_{ij}=P_j^A q_{ji}$$

或等价地，对 $i\in A$，$j\in A$

$$P_i q_{ij}=P_j q_{ji}$$

但因原链由假设是时间可逆的，所以上式成立。

例 5.6 (a) 考虑一个 $M/M/1$ 排队系统，在该系统中当来客发现系统中已有 N 人便不进去而消失。这有限容量的 $M/M/1$ 系统可看成是 $M/M/1$ 被截所得，从而是时间可逆的，极限概率为

$$P_j=\frac{(\lambda/\mu)^j}{\sum_{i=0}^{N}\left(\frac{\lambda}{\mu}\right)^i},\qquad 0\leqslant j\leqslant N$$

其中我们利用了上面例 5.5 (a) 的结果。

5.6.1 串联排队

$M/M/s$ 排队系统的时间可逆性对于排队论来说还有其它的重要含义。例如，考虑有两名服务员的系统。其中顾客依照速率为 λ 的泊松过程来到一号服务员处。在一号服务员服务完毕后，再到二号服务员前面去排队。我们假设在两个服务员前的等待场地是无限的。在任一时刻每个服务员只为一名顾客服务，i 号服务员为一名顾客服务的时间是参数为 μ_i 的指数变量，$i=1, 2$。这样一个系统称为串联系统或序贯系统（图 5.6.2)。因从一号服务员处输出的是一泊松过程，所以二号服务员面临的也是一个 $M/M/1$ 排队。然而，使用时间可逆性我们可得到更多的结果。首先我们

需要如下的引理。

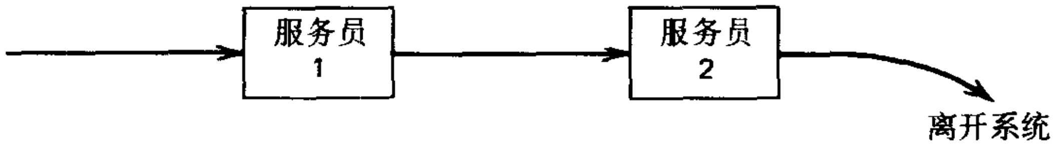


图 5.6.2 串联系统

引理 5.6.4

在一个遍历的稳态下的 $M/M/1$ 排队系统中：

(i) 目前在系统中的顾客人数与过去的诸离去时刻独立；

(ii) 一个顾客在系统中渡过的等待时间（排队等待加上服务的时间）与他离开之前的离去过程独立。

证明 (i) 因为来到过程是泊松过程，所以将来的来到的序列与目前系统中的人数独立，因此由时间可逆性知，目前在系统中的人数也必定与过去的离开的序列独立(因为时间倒看时离开被看作来到)。

(ii) 考虑一名在时刻 T_1 来到而在时刻 T_2 离去的顾客。因为系统是先来先服务且具有泊松来到，所以顾客的等待时间 T_2-T_1 与时刻 T_1 之后的来到过程独立。现在将时间倒看，我们看到，一个顾客在时刻 T_2 来到，且同一个顾客在时刻 T_1 离去(为什么是同一个顾客?)因此，由时间可逆性，看逆向过程可知，T_2-T_1 独立于(依逆方向)时刻 T_2 之后的(逆向)来到过程。但这正是时刻 T_2 之前的离去过程。

定理 5.6.5 对于稳态下的遍历串联排队：

(i) 目前在一号服务员与二号服务员处的顾客数是独立的，且

$$P\{n \text{ 人在一号服务员处，} m \text{ 人在二号服务员处}\}$$
$$=\left(\frac{\lambda}{\mu_1}\right)^n\left(1-\frac{\lambda}{\mu_1}\right)\left(\frac{\lambda}{\mu_2}\right)^m\left(1-\frac{\lambda}{\mu_2}\right)$$

(ii) 一顾客在一号服务员处的等待时间与他在二号服务员处的等待时间独立。

证明 (i) 由引理 5.6.4 (i) 知顾客们在一号服务员处的人数与过去从一号服务员处离去的诸时刻独立。由于这些过去的离开时刻构成二号服务员的来到过程，所以在此两个系统中的顾客数是独立的。联合概率公式可从独立性及例 5.5(a)中给出的 $M/M/1$ 排队系统的极限概率公式推得。

(ii) 由引理 5.6.4(ii) 可见一个指定的顾客在一号服务员处化费的时间与他离开一号服务员之前的离去过程独立，但这后一过程连同在二号服务员处的服务时间显然就决定了顾客在二号服务员处的等待时间，因此结论得证。

注记

(1) 为使定理 5.6.5(i)中的公式表示联合概率，显然必须有 $\lambda/\mu_i<1$，$i=1$，2。这是串联排队为遍历的充要条件。

(2) 虽然一指定的顾客在两名服务员处的等待时间是独立的，但有点令人惊奇的是一个顾客的排队时间是不独立的。作为一个反例，假设 λ 相对于 $\mu_1=\mu_2$ 很小；于是几乎所有的顾客在两个服务员处的排队时间为零。然而，在一个顾客在一号服务员处的排队时间为正的条件下，在二号服务员处的排队时间为正的概率至少是 $\frac{1}{2}$。（为什么?）因此，排队时间是不独立的。

5.6.2 一个随机群体模型

假设有变异的个体依照参数为 λ 的泊松过程进入一个群体。每个变种一进入就成为一个家族的始祖。在群体中一切个体独立地活动，且以指数率 ν 生育，以指数率 μ 死亡，这里我们假定 $\nu<\mu$。

以 $N_j(t)$ 记时刻 t 恰有 j 个成员($j\geqslant 0$)的家族的个数，且记

$$\underline{N}(t)=(N_1(t),N_2(t),\cdots)$$

则 $\{\underline{N}(t),t\geqslant 0\}$ 是一连续时间马尔可夫链。

对任意状态 $\underline{n}=(n_1,n_2,\cdots,n_j,\cdots)$，其中 $n_j>0$，定义状态

$$B_j\underline{n}=(n_1,n_2,\cdots,n_{j-1},n_j-1,n_{j+1}+1,\cdots),\quad j\geqslant 1$$

$$D_j\underline{n}=(n_1,n_2,\cdots,n_{j-1}+1,n_j-1,n_{j+1},\cdots),\quad j\geqslant 2$$

也定义

$$B_0\underline{n}=(n_1+1,n_2,\cdots)$$

$$D_1\underline{n}=(n_1-1,n_2,\cdots)$$

若有 j 个成员($j\geqslant 1$)的家族中有一生或一死，$B_j\underline{n}$ 与 $D_j\underline{n}$ 分别表

示从状态 $\underline{n}$ 转移到的下一状态，而 $B_0\underline{n}$ 则是出现一个变种时的下一个状态。

如果我们以 $q(\underline{n},\underline{n}')$ 记马尔可夫链的转移率，则非零的转移率仅仅是

$$q(\underline{n},B_0\underline{n})=\lambda$$

$$q(\underline{n},B_j\underline{n})=jn_j\nu,\qquad j\geqslant 1$$

$$q(\underline{n},D_j\underline{n})=jn_j\mu,\qquad j\geqslant 1$$

为了分析此马尔可夫链，值得注意家族依照一个泊松过程进入群体，且与其它家族的活动互相独立地以随机方式改变着状态。我们称一个家族处于状态 j，若它有 j 个成员。现在我们假设起初群体是空的——即 $N(0)=\underline{0}$——而且我们称一个来到的变种为 j 型的，若其家族在时刻 t 包含 j 个个体。于是从第二章命题 2.3.2 对于多个类型的推广可得 $\{N_j(t),j\geqslant 1\}$ 是独立的泊松随机变量，分别具有均值

(5.6.3) $$E[N_j(t)]=\lambda\int_0^t P_j(s)ds$$

其中 $P_j(s)$ 是时刻 s 起源的一个家族在时刻 t 包含 j 个个体的概率。

以 $P(\underline{n})$ 记极限概率，从 $N(0)=\underline{0}$ 时 $N_j(t)(j\geqslant 1)$ 是独立的泊松随机变量这一事实知，极限概率将有形式

(5.6.4) $$P(\underline{n})=\prod_{i=1}^{\infty}e^{-\alpha_i}\frac{\alpha_i^{n_i}}{n_i!}$$

$\alpha_1,\alpha_2,\cdots$ 为某些数。现在我们将决定 α_i 且同时证明此过程是时间可逆的。对形为(5.6.4)的 $P(\underline{n})$ 有

$$P(\underline{n})q(\underline{n},B_0\underline{n})=\lambda\prod_{i=1}^{\infty}e^{-\alpha_i}\frac{\alpha_i^{n_i}}{n_i!}$$

$$P(B_0\underline{n})q(B_0\underline{n},\underline{n})=(n_1+1)\mu\frac{e^{-\alpha_1}\alpha_1^{n_1+1}}{(n_1+1)!}\prod_{i=2}^{\infty}e^{-\alpha_i}\frac{\alpha_i^{n_i}}{n_i!}$$

使 $P(\underline{n})q(\underline{n},B_0\underline{n})$ 与 $P(B_0\underline{n})q(B_0\underline{n},\underline{n})$ 相等得

(5.6.5) $$\alpha_1=\lambda/\mu$$

类似的对 $j\geqslant 1$,(5.6.4)给出:

$P(\underline{n})q(\underline{n},B_j\underline{n})$将等于 $P(B_j\underline{n})q(B_j\underline{n},\underline{n})$,

若 $j\nu\alpha_j=(j+1)\mu\alpha_{j+1}$

由(5.6.5)便得

$$\alpha_j=\frac{\lambda}{j\,\nu}\left(\frac{\nu}{\mu}\right)^j$$

因此,对于(5.6.4)给出的 $P(\underline{n})$,其中 $\alpha_j=\lambda(\nu/\mu)^j/j\,\nu$,我们证明了

$$P(\underline{n})q(\underline{n},B_j\underline{n})=P(B_j(\underline{n}))q(B_j(\underline{n}),\underline{n})$$

在上式中改写 $\underline{n}$ 为 $B_{j-1}(D_j(\underline{n}))$并用上述结果,可对状态 $\underline{n}$ 与 $D_j\underline{n}$ 得到类似的结果,因此有下列

定理 5.6.6 连续时间马尔可夫链$\{N(t),t\geqslant 0\}$在稳态下是时间可逆的,其极限概率为

$$P(\underline{n})=\prod_{i=1}^{\infty}e^{-\alpha_i}\frac{\alpha_i^{n_i}}{n_i!}$$

其中

$$\alpha_i=\frac{\lambda}{i\nu}\left(\frac{\nu}{\mu}\right)^i,\qquad i\geqslant 1$$

换言之,有 i 个个体的家族的极限个数是独立的泊松随机变量,均值分别为

$$\frac{\lambda}{i\nu}\left(\frac{\nu}{\mu}\right)^j,\quad i\geqslant 1$$

α_i 除了是有 i 个成员的家族的极限平均数外还有一种有趣的解释。从(5.6.3)可见

$$E[N_i(t)]=\lambda\int_0^t q(t-s)ds$$
$$=\lambda\int_0^t q(s)ds$$

其中 $q(s)$是一个家族起源后经过时间 s 有 i 个个体的概率。

因此,

(5.6.6) $$\lim_{t\to\infty}E[N_i(t)]=\lambda\int_0^\infty q(s)ds$$

但是考虑任意一个家族且设

$$I(s)=\begin{cases}1, & \text{起源后经过时间 } s \text{ 此家族包含 } i \text{ 个成员}\\ 0, & \text{其它}\end{cases}$$

则

$$\begin{aligned}\int_0^\infty q(s)ds &=\int_0^\infty E[I(s)]ds\\ &=E[\int_0^\infty I(s)ds]\\ &=E[\text{此家族有 } i \text{ 个成员的时间}]\end{aligned}$$

因此，从(5.6.6)

$$\lim_{t\to\infty}E[N_i(t)]=\lambda E[\text{一家族有 } i \text{ 个成员的时间}]$$

且因

$$E[N_i(t)]\to\alpha_i=\frac{\lambda}{i\nu}\left(\frac{\nu}{\mu}\right)^i$$

可见

$$E[\text{一家族有 } i \text{ 个成员的时间}]=\frac{(\nu/\mu)^i}{i\nu}$$

今考虑稳态下的群体模型且假设目前的状态是$\underline{n}^*$。我们想确定一个给定的有 i 个成员的家族是群体中最老(意即起源最早)的家族的概率。似乎我们可能用过程的时间可逆性去推断出，此概率与该家族是目前存活的家族中最后起源的概率相同。然而，不幸的是此结论不能立即得到。因此就我们的状态空间，不可能由观察全过程去决定一指定的家族灭绝的确切时间。因此我们需要一个提供更多信息的状态空间——一个使我们有可能去跟踪一个指定的家族在整个时间中的演变的状态空间。

出于技术上的原因，从不允许多于 M 个家族存在的截尾模型着手，$M\geqslant\Sigma_i n_i^*$，将是最简易的。即无论何时在群体中若存在 M 个家族，则不允许任何其它的变种进入。注意到，由命题 5.6.3 具有状态 $\underline{n}$ 的截尾过程仍是时间可逆的且具有平稳分布。

$$P(\underline{n}) = C\prod_{i=1}^{\infty} e^{-\alpha_i}\frac{\alpha_i^{n_i}}{n_i!}, \qquad \sum_{i=1}^{\infty} n_i \leqslant M$$

其中 $\alpha_i = \lambda(\nu/\mu)^i/i\nu$。

为了随时间前进跟住指定的家族的行踪，我们不得不将不同的家族标明。让我们使用标记 1，2…，M 且约定每当一个新的家族起源(即一个变种出现)时，它的标记从那时尚未被取用的标记集合中等可能地选取。若以 s_i 记标记为 i，$i=1$，2，…，M，的家族的个体数（$s_i=0$ 意指目前标记为 i 的家族不存在），则我们可认为此过程具有状态 $\underline{s}=(s_1,\ \cdots,\ s_M)$，$s_i\geqslant 0$。对于给定的 $\underline{s}$，令 $\underline{n}(\underline{s})=(n_1(\underline{s}),\cdots,n_k(\underline{s}),\cdots)$，其中 $n_i(\underline{s})$ 如以前一样，是有 i 个成员的家族的个数。即

$$n_i(\underline{s}) = \text{使 } s_j=i \text{ 的 } j \text{ 的个数。}$$

为了得到具有状态 $\underline{s}$ 的马尔可夫链的平稳分布，注意到对 $n=\underline{n}(\underline{s})$

$$\begin{aligned} P(\underline{s}) &= P(\underline{n})P(\underline{s}\,|\,\underline{n}) \\ &= P(\underline{s}\,|\,\underline{n})C\prod_{i=1}^{\infty} e^{-\alpha_i}\frac{\alpha_i^{n_i}}{n_i!} \end{aligned}$$

由于一切标记是随机地选取的，所以直观上对给定的向量 $\underline{n}$，与它相一致的全部

$$\frac{M!}{(M-\sum_{i=1}^{\infty} n_i)!\ \prod_{i=1}^{\infty} n_i!}$$

个可能的向量 $\underline{s}$ 是等可能发生的(即，如果 $M=3$，$n_1=n_2=1$，$n_i=0, i\geqslant 3$，则存在两个家族(一个有一个成员与一个有二个成员)，且直观上看来，六个可能的状态 $\underline{s}$（0，1，2 的全部排列)是等可能的)。因此直观上看来

(5.6.7) $$P(\underline{s}) = \frac{(M-\sum_{i=1}^{\infty} n_i)!\ \prod_{i=1}^{\infty} n_i!}{M!}C\prod_{i=1}^{\infty} e^{-\alpha_i}\frac{\alpha_i^{n_i}}{n_i!}$$

其中 $n_i = n_i(\underline{s})$ 而 $\alpha_i = \lambda\ (\nu/\mu)^i/i\ \nu$。现在我们将验证上面的公式且

同时证明具有状态 $\underline{s}$ 的链是时间可逆的。

命题 5.6.7

具有状态 $\underline{s}=(s_1,\cdots,s_M)$ 的链是时间可逆的且具有（5.6.7）所给出的平稳分布。

证明 对于一个向量 $\underline{s}=(s_1,\cdots,s_i\cdots,s_M)$，令

$$B_i(\underline{s})=(s_1,\cdots,s_i+1,\cdots,s_M)$$

即是若标记为 i 的家族的一位成员生一后代时则 $B_i(\underline{s})$ 是继 $\underline{s}$ 之后的状态，现在对 $s_i>0$

$$q(\underline{s},B_i(\underline{s}))=s_i\nu,\qquad s_i>0$$

$$q(B_i(\underline{s}),\underline{s})=(s_i+1)\mu,\qquad s_i>0$$

若

$$n(\underline{s})=(n_1,\cdots,n_{s_i},n_{s_i+1},\cdots)$$

则也有

$$n(B_i(\underline{s}))=(n_1,\cdots,n_{s_i}-1,n_{s_i+1}+1,\cdots)$$

因此对于由（5.6.7）所给出的 $P(\underline{s})$ 及 $s_i>0$，

(5.6.8) $$P(\underline{s})q(\underline{s},B_i(\underline{s}))=P(B_i(\underline{s}))q(B_i(\underline{s}),\underline{s})$$

等价于

$$\frac{n_{s_i}\alpha_{s_i}}{n_{s_i}!\ n_{s_i+1}!}s_i\nu=\frac{(n_{s_i+1}+1)\ \alpha_{s_i+1}}{(n_{s_i}-1)!\ (n_{s_i+1}+1)!}(s_i+1)\mu$$

或

$$\alpha_{s_i}s_i\nu=\alpha_{s_i+1}(s_i+1)\mu$$

或，因为 $\alpha_i=\lambda(\nu/\mu)^i/i\nu$，

$$\lambda\left(\frac{\nu}{\mu}\right)^{s_i}=\lambda\left(\frac{\nu}{\mu}\right)^{s_i+1}\left(\frac{\mu}{\nu}\right)$$

此式显然成立。

因为状态是 $\underline{s}$ 时有 $M-\sum_{i=1}^{\infty}n_i(\underline{s})$ 个标记可用于一新诞生的变种，所以可见

$$q(\underline{s},B_i(\underline{s})=\frac{\lambda}{M-\sum_{i=1}^{\infty}n_i(\underline{s})},\qquad \text{若 } s_i=0$$

$$q(B_i(\underline{s}),\underline{s})=\mu,\qquad \text{若 } s_i=0$$

容易证明方程(5.6.8)在这种情形下也成立，于是证毕。

系 5.6.8

若在稳态下成员为 i 个的家族有 n_i 个，$i>0$，则一个指定的有 i 个成员的家族是此群体中最老的家族的概率是 $i/\sum_j jn_j$。

证明 考虑具有状态 $\underline{s}$ 的截尾过程且假设状态 $\underline{s}$ 满足 $n_i(\underline{s})=n_i$。一个给定的有 i 个成员的家族是最老的，如果时间倒过去看它存在得最久。但由时间可逆性，时间倒过去的过程与时间向前看的过程有同样的概率分布，因而指定的家族是最老的概率就是此家族是目前群体中的所有家族之中将存活最久的概率。但对于各个个体来说，其子孙是存活时间最久的概率是同样的，而在群体中有 $\sum_j jn_j$ 个个体，其中 i 个属于此家族，故这种情形下结论成立，令 $M\to\infty$ 则得一般情形下的证明。

注记 我们选择截尾过程处理，是因为便于猜测具有状态 $\underline{s}$ 的标记过程的极限概率。

5.7 逆向链对排队论的应用

逆向链是一个十分有用的概念，即使过程不是时间可逆的。为了明瞭这一点，我们从下列与第四章定理 4.7.2 相类似的连续时间情形的结果开始。

定理 5.7.1 以 q_{ij} 记一不可约连续时间马尔可夫链的转移率。如果我们能找到一组数 q_{ij}^*，$i,j\geqslant 0,i\neq j$，及一组和为 1 的非负数 P_i，$i\geqslant 0$，使得

$$P_iq_{ij}=P_jq_{ji}^*,\qquad i\neq j$$

且

$$\sum_{j\neq i}q_{ij}=\sum_{j\neq i}q_{ij}^*,\qquad i\geqslant 0$$

则 q_{ij}^* 是逆向链的转移率而 P_i 是(两个链的)极限概率。

定理 5.7.1 的证明留为练习。

如果我们能够猜到逆向链与极限概率，则可用定理 5.7.1 证

实我们的猜测。为说明这种方法，考虑下列排队网络的模型，它实质上推广了前节的串联排队模型。

5.7.1 排队网络

考虑一个有 k 个服务员的系统，顾客从系统外面来到各个服务员处，i 号服务员处的来到过程($i=1,\cdots,k$)是独立的速率为 r_i 的泊松过程；顾客在 i 号处排队直至轮到为他们服务。每当一个顾客被 i 号服务员服务完毕，他以概率 P_{ij} 到 j 号服务员处参加排队，这里 $\sum_{j=1}^{k} P_{ij}\leqslant 1$，而 $1-\sum_{j=1}^{k} P_{ij}$ 表示一顾客经 i 号服务员服务完毕后离开系统的概率。

若以 λ_j 记顾客来到 j 号服务员处的总来到速率，则 λ_j 可作为

(5.7.1) $$\lambda_j=r_j+\sum_{i=1}^{k}\lambda_i P_{ij},\qquad i=1,\cdots,k$$

的解而得到。这是因为 r_j 是顾客从系统之外来到 i 号处的速率，而 λ_i 是顾客离开 i 号服务员的速率(进的速率必须等于出的速率)，$\lambda_i P_{ij}$ 是自 i 号服务员处到 j 号处的速率。

此模型可作为具有状态($n_1,n_2,\cdots,n_k$)的连续时间马尔可夫链，其中 n_i 表示在 i 号服务员处的顾客数，根据串联排队的结论我们可以期望在各个服务员处的顾客数是独立的随机变量。把极限概率记为 $P(n_1,n_2,\cdots,n_k)$，我们先试着证明

(5.7.2) $$P(n_1,n_2,\cdots,n_k)=P_1(n_1)P_2(n_2)\cdots P_k(n_k)$$

这样，这里的 $P_i(n_i)$ 就是在 i 号服务员处有 n_i 个顾客的极限概率。为了证明这些概率确实有上面的形式并得到 $P_i(n_i)$，$i=1,\cdots,k$，我们要先离开本题而对逆向过程作推测。

在逆向过程中，当一顾客离开 i 号服务员时，以某概率去 j 号处，此概率可望不依赖于过去。若此概率(记作 $\overline{P}_{ij}$)确实不依赖于过去，那么其值是多少呢？要回答这个问题，先注意来到一个服务员处的速率必须等于离开他的速率，从而在正向及逆向过程中在 j 号服务员处的来到速率都是 λ_j。因为在正向过程中，顾客从 j

到 i 转移率必等于他们在逆向过程中从 i 到 j 的转移率，这意味着

$$\lambda_j P_{ji} = \lambda_i \overline{P}_{ij}$$

或

$$\overline{P}_{ij} = \frac{\lambda_j P_{ji}}{\lambda_i} \tag{5.7.3}$$

于是我们可望在逆向过程中一名顾客离开 i 号服务员时以概率 $\overline{P}_{ij}=\lambda_j P_{ji}/\lambda_i$ 到 j 号服务员处。

在逆向过程中从系统外来到 i 号服务员处对应于正向过程中从 i 号处离开系统，因而以速率 $\lambda_i(1-\sum_{j=1}^{k} P_{ij})$ 发生。最好不过的可能性是它为一泊松过程；于是我们作下列猜测。

猜测　逆向随机过程与原过程是同类型的网络过程。从系统之外到 i 号服务员处的来到是速率为 $\lambda_i(1-\sum_{j=1}^{k} P_{ij})$ 的泊松过程，且从 i 离去到 j 的概率 $\overline{P}_{ij}$ 由(5.7.3)给出。i 号的服务率是指数率 μ_i。此外，极限概率满足

$$P(n_1,n_2,\cdots,n_k) = P_1(n_1)P_2(n_2)\cdots P_k(n_k)$$

为了证明此推测并得到 $P_i(n_i)$，先考虑一名外来顾客引起的转移，即考虑状态 $\underline{n}=(n_1, \cdots, n_i, \cdots, n_k)$ 与 $\underline{n}'=(n_1, \cdots, n_i+1, \cdots, n_k)$。既然

$$q_{n,n'} = r_i$$

且若猜测为真，应有

$$\begin{aligned} q^*_{n',n} &= \mu_i(1-\sum_{j=1}^{k}\overline{P}_{ij}) \\ &= \mu_i \frac{(\lambda_i - \sum_{j=1}^{k}\lambda_j P_{ji})}{\lambda_i} \\ &= \frac{\mu_i r_i}{\lambda_i} \qquad \text{（由(5.7.1)）} \end{aligned}$$

而

$$P(\underline{n}) = \prod_{j=1}^{k} P_j(n_j), \quad P(\underline{n}') = P_i(n_i + 1)\prod_{j\neq i} P_j(n_j)$$

因此从定理 5.7.1 我们需要

$$r_i \prod_{j=1}^{k} P_j(n_j) = \frac{\mu_i r_i}{\lambda_i} P_i(n_i + 1) \prod_{j\neq i} P_j(n_j)$$

或

$$P_i(n_i + 1) = \frac{\lambda_i}{\mu_i} P_i(n_i)$$

即

$$P_i(n + 1) = \frac{\lambda_i}{\mu_i} P_i(n) = \left(\frac{\lambda_i}{\mu_i}\right)^2 P_i(n - 1) \cdots = \left(\frac{\lambda_i}{\mu_i}\right)^{n+1} P_i(0)$$

利用 $\sum_{n=0}^{\infty} P_i(n) = 1$ 得

$$P_i(n) = \left(\frac{\lambda_i}{\mu_i}\right)^n \left(1 - \frac{\lambda_i}{\mu_i}\right) \tag{5.7.4}$$

于是为使猜测成立，λ_i/μ_i 必须小于 1 且 P_i 必须如上式所示。

为了继续证明此猜测，考虑从 j 号服务员处离去到 i 号服务员处引起的转移。即令 $\underline{n} = (n_1, \cdots, n_i, \cdots, n_j, \cdots, n_k)$ 及 $\underline{n}' = (n_1, \cdots, n_i + 1, \cdots, n_j - 1, \cdots, n_k)$，其中 $n_j > 0$。既然

$$q_{\underline{n},\underline{n}'} = \mu_j P_{ji}$$

由猜测可得

$$q^*_{\underline{n}',\underline{n}} = \mu_i \overline{P}_{ij}$$

我们需要证明

$$P(\underline{n}) \mu_j P_{ji} = P(\underline{n}') \mu_i \overline{P}_{ij}$$

利用(5.7.4)也就是要证

$$\lambda_j P_{ji} = \lambda_i \overline{P}_{ij}$$

而这正是 $\overline{P}_{ij}$ 的定义。

由于定理 (5.7.1) 所需的其余事实的验证可用同样的方法进行，所以我们证明了下列

定理 5.7.2 假设对一切 i 有 $\lambda_i < \mu_i$，在稳态下，在 i 号服务

员处的顾客数量是独立的且极限概率为

$$P(n_1,\cdots,n_k)=\prod_{i=1}^{k}\left(\frac{\lambda_i}{\mu_i}\right)^{n_i}\left(1-\frac{\lambda_i}{\mu_i}\right)$$

从逆向链我们也有下列的

系 5.7.3

从 i 号服务员处离开系统的顾客过程是独立的泊松过程，速率分别为 $\lambda_i(1-\sum_{j=1}^{k}P_{ij})$，$i=1$，…，$k$。

证明 我们已证明在逆向过程中顾客依照速率为 $\lambda_i(1-\sum_{j=1}^{k}P_{ij})$ 的独立泊松过程从系统之外来到 i 号服务员处。由于在逆向过程中从外面来到 i 号服务员处对应于正向过程中从 i 号服务员离开系统，从而结论得证。

注记

（1）体现在定理 5.7.2 中的这个结果是颇为惊人的。因为它说在 i 号服务员处的顾客数的分布与具有参数 λ_i 与 μ_i 的一个 $M/M/1$ 系统的分布相同。惊人的地方在于在网络模型中在节点 i 处的来到过程不必是泊松过程。因为如果存在一顾客多次光顾一位服务员的可能性(称为反馈的情形)，则来到过程不是泊松过程。一个易于说明此事的例子是一个单服务员系统其服务效率（即指数服务时间的参数）相对于外来的甚小的来到速率显得非常大，又假设刚服务完毕的顾客以概率 $p=0.9$ 反馈到系统中。因此在一来到时刻有很大的概率在短时间后又有一个来到(即反馈来到)；在任意时刻只有很少的机会不久就发生一个来到(因为 λ 很小)。因此这个来到过程不具有独立增量从而不可能是泊松过程。

（2）此模型可推广为各个服务站是多个服务员的系统(即 i 号服务员工作情况如同 $M/M/k_i$ 系统而不是 $M/M/1$ 系统)。在各个服务站的顾客的极限数仍将是独立的，且在一个服务员处的顾客数的极限分布如同来到过程是泊松过程时一样。

5.7.2 爱尔朗消失公式

考虑一消失排队模型，其中顾客们按照速率为 λ 的泊松过程

来到一个有 k 位服务员的系统。所谓是一个消失模型，意指任何来客发现所有服务员都忙着便不进入系统，而自系统中消失了。服务员的服务时间假定具有分布 G。我们将假设 G 是一个具有密度 g 和失效率函数 $\lambda(t)$ 的连续分布。即不严格地说 $\lambda(t)=g(t)/\overline{G}(t)$ 是已经历 t 单位时间之久的服务行将结束的瞬时概率强度。

在分析上面的系统时令任意时刻的状态为该时刻在服务中的顾客的顺序年龄，即状态是 $\underline{x}=(x_1, x_2\cdots, x_n)$，$x_1\leqslant x_2\leqslant\cdots\leqslant x_n$，如果有 n 个顾客在接受服务，最近来到的一个是 x_1 个单位时间前来到的，次最近的是 x_2 个单位时间前来到的，等等。相继的状态形成的过程是一个马尔可夫过程，在给定现在及过去的状态时任何将来状态的条件分布只依赖于现在的状态，尽管这过程不是一个连续时间的马尔可夫链，但我们对链所发展的理论能扩展使用于这个过程，我们将在此基础上分析这个模型。

我们将试用逆向过程去得到极限概率密度 $p(x_1, x_2, \cdots, x_n)$，$1\leqslant n\leqslant k$，$x_1\leqslant x_2\leqslant\cdots\leqslant x_n$，以及系统是空的极限概率 $P(\phi)$。由于一位在服务中的顾客的年龄从 0（他刚来到）线性地增加到他的服务时间（他即离去），所以如果我们倒过来看，将看到的是他的剩余服务时间。因为在本系统中永远不会多于 k 个顾客，所以我们作如下的猜想。

猜测　逆向过程也是一个 k 个服务员的消失系统，服务时间分布为 G，来到依照速率为 λ 的泊松过程发生。在任何时刻的状态表示当时在服务中的顾客们的顺序剩余服务时间。

现在我们试着去证明上面的猜测且同时得到极限分布。对于任何状态 $\underline{x}=(x_1, x_i\cdots, x_n)$ 令 $e_i(\underline{x})=(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$。现在在原过程中当状态是 $\underline{x}$ 时，它瞬时间转到 $e_i(\underline{x})$ 的概率密度等于 $\lambda(x_i)$，因为已服务了 x_i 时间的人必须瞬时间结束他的服务。类似地，在逆向过程中如果状态是 $e_i(\underline{x})$，它瞬间到 $\underline{x}$，如果一个有 x_i 服务时间的顾客瞬间来到。由此可见

在正向过程中：$\underline{x}\rightarrow e_i(\underline{x})$ 以概率强度 $\lambda(x_i)$；

在逆向过程中：$e_i(\underline{x}) \to \underline{x}$ 以（联合）概率强度 $\lambda g(x_i)$。因此如果 $p(\underline{x})$表示极限密度，则按照定理 5.7.1 我们将需要

$$p(\underline{x})\lambda(x_i) = p(e_i(\underline{x}))\lambda g(x_i)$$

或，因为 $\lambda(x_i)=g(x_i)/\overline{G}(x_i)$，

$$p(\underline{x}) = p(e_i(\underline{x}))\lambda\overline{G}(x_i)$$

令 $i=1$ 并叠代上式得

(5.7.5)
$$\begin{aligned} p(\underline{x}) &= \lambda\overline{G}(x_1)p(e_1(\underline{x})) \\ &= \lambda\overline{G}(x_1)\lambda\overline{G}(x_2)p(e_2(e_1(\underline{x}))) \\ &\vdots \\ &= \prod_{i=1}^{n}\lambda\overline{G}(x_i)P(\phi) \end{aligned}$$

在全体向量$\underline{x}$ 上积分得

(5.7.6)

$$\begin{aligned} P\{n \text{ 人在系统中}\} &= P(\phi)\lambda^n \idotsint\limits_{x_1\leqslant x_2\leqslant\cdots\leqslant x_n} \prod_{i=1}^{n}\overline{G}(x_i)dx_1dx_2\cdots dx_n \\ &= P(\phi)\frac{\lambda^n}{n!}\idotsint\limits_{x_1,x_2,\cdots,x_n}\prod_{i=1}^{n}\overline{G}(x_i)dx_1dx_2\cdots dx_n \\ &= P(\phi)\frac{(\lambda E[S])^n}{n!}, \quad n=1,2,\cdots k \end{aligned}$$

其中 $E[S]=\int\overline{G}(x)dx$ 是平均服务时间，且只要利用

$$P(\phi) + \sum_{n=1}^{k}P\{n \text{ 人在系统中}\} = 1$$

我们得到

(5.7.7)
$$P\{n \text{ 人在系统中}\} = \frac{\dfrac{(\lambda E[S])^n}{n!}}{\displaystyle\sum_{i=0}^{k}\frac{(\lambda E[S])^i}{i!}}, \quad n=0,1,\cdots,k$$

从(5.7.5)我们有

(5.7.8)
$$p(\underline{x})=\frac{\lambda^n\prod_{i=1}^{n}\overline{G}(x_i)}{\sum_{i=0}^{k}\frac{(\lambda E[S])^i}{i!}}$$

并看到在系统中有 n 个人的条件下顺序年龄的条件分布密度为

$$p\{\underline{x}|n\text{ 人在系统中}\}=\frac{p(\underline{x})}{P\{n\text{ 人在系统中}\}}$$
$$=n!\prod_{i=1}^{n}\frac{\overline{G}(x_i)}{E[S]}$$

由于 $\overline{G}(x)/E[S]$ 正是 G 的平衡分布 G_e 的密度，我们看到，如果此猜测是对的，则系统中人数的极限分布仅通过其均值依赖于 G 且由(5.7.6)给出，且在系统中有 n 人的条件下它们的(无序)年龄独立并同分布于 G 的平衡分布 G_e。

为了完成此猜测的证明，必须考虑正向过程在 $n<k$ 时从 $\underline{x}$ 到 $(0,\ \underline{x})=(0,\ x_1,\ x_2,\ \cdots,\ x_n)$ 的转移。现在

在正向过程中：$\underline{x}\to(0,\ \underline{x})$ 以瞬时强度 λ；

在逆向过程中：$(0,\ \underline{x})\to\underline{x}$ 以概率 1。

因此联系定理 5.7.1 我们必须验证

$$p(\underline{x})\lambda=p(0,\underline{x})$$

它得自(5.7.8)，因为 $\overline{G}(0)=1$。

所以，假定(5.7.1)的类似结论成立，我们已经证明了下列

定理 5.7.4 在系统中的顾客人数的极限分布为

(5.7.9)
$$P\{n\text{ 人在系统中}\}=\frac{\frac{(\lambda E[S])^n}{n!}}{\sum_{i=0}^{k}\frac{(\lambda E[S])^i}{i!}},\qquad n=0,1,\cdots,k$$

且在系统中有 n 人的条件下这 n 人的年龄(或剩余时间)独立并同分布于 G 的平衡分布。

所考虑的模型常称之为爱尔朗消失系统，而等式(5.7.9)称为

爱尔朗消失公式。

利用逆向过程也可得如下的系。

系 5.7.5

在爱尔朗消失模型中离去过程(既包括服务完毕的也包括消失的顾客)是速率为 λ 的泊松过程。

证明 因为在逆向过程中所有顾客(包括消失的)的来到构成一泊松过程，故得上述结论。

5.7.3 $M/G/1$ 共用加工系统

假设顾客按照速率为 λ 的泊松过程来到，每个顾客需要服务员做加工工作，工作量是随机的，其分布为 G。服务员以每单位时间做一单位工作量的速率加工，且在系统中的全体顾客中间等分他的时间，也即每当系统中有 n 个顾客时每人都以每单位时间做 $1/n$ 单位工作量的速率得到服务。

以 $\lambda(t)=g(t)/\overline{G}(t)$ 记服务分布的失效率函数，且假设 $\lambda E[S]<1$，其中 $E[S]$ 是 G 的均值。

为了分析此模型，令任意时刻的状态为仍在系统中的顾客已完成的工作量的顺序向量，也即状态是 $\underline{x}=(x_1, x_2, \cdots, x_n)$，$x_1\leqslant x_2\leqslant\cdots\leqslant x_n$，如果系统中有 n 个顾客而 $x_1, \cdots, x_n$ 是这 n 个顾客已完成的工作量。以 $p(\underline{x})$ 与 $P(\phi)$ 记极限概率密度与系统为空的极限概率。关于逆向过程我们作如下的猜测。

猜测 逆向过程是同一类型的系统，顾客依照速率为 λ 的泊松过程来到，要求加工的工作量分布为 G，其状态表示系统中的顾客的顺序剩余工作量。

为了验证以上猜测且同时得到极限分布，令 $e_i(\underline{x})=(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ 若 $\underline{x}=(x_1, \cdots, x_n)$，$x_1\leqslant x_2\leqslant\cdots\leqslant x_n$。注意到

在正向过程中：$\underline{x}\to e_i(\underline{x})$ 以概率强度 $\lambda(x_i)/n$；

在逆向过程中：$e_i(\underline{x})\to\underline{x}$ 以(联合)概率强度 $\lambda g(x_i)$。

这些可以用与前一节同样的方法得到，不同的是若在系统中有 n

个人则一个已完成工作量 x_i 的顾客将立即加工完毕的概率强度为 $\lambda(x_i)/n$。

因此，若 $p(x)$ 是极限密度，则依定理 5.7.1 我们需要

$$p(\underline{x})\frac{\lambda(x_i)}{n}=p(e_i(\underline{x}))\lambda g(x_i)$$

或等价地

$$\begin{aligned}(5.7.10)\qquad p(\underline{x})&=n\overline{G}(x_i)p(e_i(\underline{x}))\lambda\\&=n\overline{G}(x_i)(n-1)\overline{G}(x_j)p(e_j(e_i(\underline{x})))\lambda^2,\quad j\neq i\\&\vdots\\&=n!\ \lambda^n P(\phi)\prod_{i=1}^{n}\overline{G}(x_i)\end{aligned}$$

如在(5.7.6)中那样，在全体向量$\underline{x}$ 上积分得

$$P\{n\text{ 人在系统中}\}=(\lambda E[S])^n P(\phi)$$

利用

$$P(\phi)+\sum_{n=1}^{\infty}P\{n\text{ 人在系统中}\}=1$$

给出

$$P\{n\text{ 人在系统中}\}=(\lambda E[S])^n(1-\lambda E[S]),\qquad n\geqslant 0$$

从(5.7.10)也得，在系统中有 n 人的条件下，顺序的已完成的工作量的条件分布密度为

$$\begin{aligned}p(\underline{x}|n)&=p(\underline{x})/P\{n\text{ 人在系统中}\}\\&=n!\prod_{i=1}^{n}\frac{\overline{G}(x_i)}{E[S]}\end{aligned}$$

即在系统中有 n 人的条件下无顺序的已完成的工作量是独立的且服从于 G 的平衡分布 G_e。

以上都是以假设此猜测成立为基础的。为了完成猜测成立的证明我们必须验证

$$p(\underline{x})\lambda=p(0,\underline{x})\frac{1}{n+1}$$

因为处于状态 $(\varepsilon,\underline{x})$ 的逆向过程在时间 $(n+1)\varepsilon$ 后到状态$\underline{x}$，所

以有上列有关等式。由于上式很容易得到验证，于是我们证明了下列

定理 5.7.6　对共用加工模型，系统中的顾客人数具有分布

$$P\{n\text{ 人在系统中}\} = (\lambda E[S])^n(1-\lambda E[S]), \qquad n \geqslant 0$$

在系统中有 n 人的条件下，已完成的(或剩余的)工作量是独立的且具有分布 G_e。离去过程是速率为 λ 的泊松过程。

若我们以 L 记系统中的平均人数，而 W 记一顾客在系统中渡过的平均时间，则

$$L = \sum_{n=0}^{\infty} n(\lambda E[S])^n(1-\lambda E[S])$$
$$= \frac{\lambda E[S]}{1-\lambda E[S]}$$

从公式 $L=\lambda W$(见第三章的 3.6.1 节)我们能得到 W，故有

$$W = \frac{L}{\lambda} = \frac{E[S]}{1-\lambda E[S]}$$

注记　非常有趣的是一顾客在系统中渡过的平均时间仅仅通过 G 的均值而依赖于 G。例如，考虑两个这样的系统：在一个系统中工作量恒为 1，而另一个系统中工作量都服从均值为 1 的指数分布。从定理 5.7.6 知，一位来客所见到的系统中人数的分布对两个系统是同样的。然而，在前一系统中一位来客发现顾客的剩余工作量服从(0,1)上均匀分布(这就是均值为 1 的决定性随机变量分布的平衡分布)，而在第二个系统中剩余工作量都是均值为 1 的指数分布。因此一位来客在两个系统中的平均时间相同是十分惊奇的。当然一个顾客在系统中渡过时间的分布，与此分布的均值相反，依赖于整个分布 G，而不是只依赖于它的均值。

此模型中另一个有趣的事情是计算已知来客的工作量为 y 的条件下他在系统中渡过的条件平均时间。为计算此量，固定 y 且说一顾客是“特殊的”，若他的工作量在 y 与 $y+\varepsilon$ 之间。由 $L=\lambda W$ 我们有

系统中特殊顾客的平均数＝(特殊顾客的平均来到速度)

$\times$（一个特殊顾客在系统中渡过的平均时间）。

为了决定系统中特殊顾客的平均人数，让我们首先确定在系统中的任一顾客的总工作量的密度。假设这样一位顾客已经受到加工的工作量为 x，则顾客的工作量的条件密度是

$$f(w \mid 已受加工\ x) = g(w)/\overline{G}(x), \qquad x \leqslant w$$

但从定理 5.7.6 知，系统中任一顾客已接受加工的工作量具有分布 G_e。因此在系统中的某人的总工作量的密度是

$$\begin{aligned} f(w) &= \int_0^w \frac{g(w)}{\overline{G}(x)} dG_e(x) \\ &= \int_0^w \frac{g(w)}{E[S]} dx \qquad \left(因\ dG_e(x) = \frac{\overline{G}(x)}{E[S]} dx\right) \\ &= wg(w)/E[S] \end{aligned}$$

因此系统中特殊顾客的平均数是

$$\begin{aligned} &E[系统中工作量在\ y\ 与\ y+\varepsilon\ 之间的人数] \\ &\quad = Lf(y)\varepsilon + o(\varepsilon)) \\ &\quad = Lyg(y)\varepsilon/E[S] + o(\varepsilon)) \end{aligned}$$

另外，工作量在 y 与 $y+\varepsilon$ 之间的顾客的平均来到速率是

$$平均来到速率 = \lambda g(y)\varepsilon + o(\varepsilon)$$

由此可见

$$\begin{aligned} &E[在系统中的时间 \mid 工作量在(y, y+\varepsilon)\ 中] \\ &\quad = \frac{g(y)\varepsilon}{\lambda g(y)\varepsilon} \frac{Ly}{E[S]} + \frac{o(\varepsilon)}{\varepsilon} \end{aligned}$$

令 $\varepsilon \to 0$ 我们得到

$$(5.7.11) \qquad \begin{aligned} E[在系统中的时间 \mid 工作量是\ y] &= \frac{y}{\lambda E[S]} L \\ &= \frac{y}{1-\lambda E[S]} \end{aligned}$$

所以一位需要加工 y 单位工作量的顾客在系统中的平均时间也仅仅通过其均值而依赖分布 G。作为验证上式的一种方法，注意

$$\begin{aligned} W &= E[在系统中的时间] \\ &= \int E[在系统中的时间 \mid 工作量是\ y] dG(y) \end{aligned}$$

$$=\frac{E[S]}{1-\lambda E[S]} \qquad (由 (5.7.11))$$

与前面所得相同。

5.8 一致化

考虑一个连续时间马尔可夫链，它在一个状态中渡过的平均时间对所有状态都是同样的，也即假设对一切状态 i，$\nu_i=\nu$。此时由于对每个状态一次逗留的时间服从参数为 ν 的指数分布，所以如果我们以 $N(t)$ 记到时间 t 为止状态转移的次数，则 $\{N(t), t \geqslant 0\}$ 是一个速率为 ν 的泊松过程。

为了计算转移概率 $P_{ij}(t)$，我们可对 $N(t)$ 取条件如下：

$$P_{ij}(t)=P\{X(t)=j|X(0)=i\}$$
$$=\sum_{n=0}^{\infty}P\{X(t)=j|X(0)=i,N(t)=n\}P\{N(t)=n|X(0)=i\}$$
$$=\sum_{n=0}^{\infty}P\{X(t)=j|X(0)=i,N(t)=n\}e^{-\nu t}\frac{(\nu t)^n}{n!}$$

到时间 t 为止已有 n 次转移这个事实给了我们一些关于在所到达的前 n 个状态的逗留时间的信息，但由于在各状态逗留的时间的分布对全体状态都相同，所以并没有给我们有关到达哪些状态的信息。因此

$$P\{X(t)=j|X(0)=i,N(t)=n\}=P_{ij}^n$$

其中 P_{ij}^n 正是具有转移概率 P_{ij} 的离散时间马尔可夫链的 n 阶转移概率；所以当 $\nu_i\equiv\nu$ 时，有

$$P_{ij}(t)=\sum_{n=0}^{\infty}P_{ij}^n e^{-\nu t}\frac{(\nu t)^n}{n!} \tag{5.8.1}$$

从计算观点上看上面的方程十分有用，因为它使我们能取一个部份和并计算(用转移概率矩阵乘法)有关的 n 步转移概率 P_{ij}^n 去逼近 $P_{ij}(t)$。

虽然方程(5.8.1)的应用范围显得十分有限，因为要假设

$\nu_i \equiv \nu$，但利用一种允许一个状态到它自身的虚转移的技巧，多数马尔可夫链能被放进这种形式。为了明瞭怎样作法，考虑任意的 ν_i 有界的马尔可夫链，设 ν 是满足下式的任何数

$$(5.8.2) \qquad \nu_i \leqslant \nu, \qquad \text{对一切 } i$$

当过程处于状态 i 时实际上以速率 ν_i 离去；而这等价于假设转移以速率 ν 发生，但只有 ν_i/ν 部分是真正转移出 i，而其余的部分是虚转移，过程仍在状态 i。换言之，任何满足条件(5.8.2)的马尔可夫链可看作是这样一个过程，它在状态 i 渡过一个参数为 ν 的指数时间而后以概率 P_{ij}^* 转移到 j，其中

$$(5.8.3) \qquad P_{ij}^* = \begin{cases} 1 - \dfrac{\nu_i}{\nu}, & j = i \\ \dfrac{\nu_i}{\nu} P_{ij}, & j \neq i \end{cases}$$

因此，从(5.8.1)转移概率可算得为

$$P_{ij}(t) = \sum_{n=0}^{\infty} P_{ij}^{*n} e^{-\nu t} \frac{(\nu t)^n}{n!}$$

其中 P_{ij}^{*n} 是对应于(5.8.3)的 n 步转移概率。这种引进一个状态到它自身的转移从而使自每个状态转移出去的速率同一的技巧，被称为一致化。

例 5.8 (a)　让我们考虑例 5.4(a) 的两状态链，它有

$$P_{01} = P_{10} = 1$$

$$\nu_0 = \lambda, \qquad \nu_1 = \mu$$

令 $\nu = \lambda + \mu$，上述链的一致化是把它考虑成一个具有

$$P_{00} = \frac{\mu}{\lambda + \mu} = 1 - P_{01}$$

$$P_{10} = \frac{\mu}{\lambda + \mu} = 1 - P_{11}$$

$$\nu_i = \lambda + \mu, \quad i = 1, 2$$

的连续时间马尔可夫链。

由于 $P_{00} = P_{10}$，进入状态 0 的转移概率是 $\mu/(\lambda + \mu)$，与目前的状态无关。由于对于状态 1 有类似的结果，得 n 步转移概率为

$$P_{10}^n = \frac{\mu}{\lambda+\mu}, \qquad n \geqslant 1, \qquad i = 0,1$$

因此，

$$P_{00}(t) = \sum_{n=0}^{\infty} P_{00}^n e^{-(\lambda+\mu)t} \frac{[(\lambda+\mu)t]^n}{n!}$$

$$= e^{-(\lambda+\mu)t} + [1 - e^{-(\lambda+\mu)t}] \frac{\mu}{\lambda+\mu}$$

$$= \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}$$

类似地，

$$P_{11}(t) = \sum_{n=0}^{\infty} P_{11}^n e^{-(\lambda+\mu)t} \frac{[(\lambda+\mu)t]^n}{n!}$$

$$= \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t}$$

另一有趣的量是到时刻 t 为止过程处在一指定状态的总时间。也即，若 $X(t)$ 记时刻 t 的状态（0 或 1）。令 $S_i(t), i = 1,0$，为

(5.8.4)
$$S_1(t) = \int_0^t X(s)ds$$

$$S_0(t) = \int_0^t (1 - X(s))ds$$

则 $S_i(t)$ 表示到时刻 t 为止在状态 i 的逗留时间。此随机过程 $\{S_i(t), t > 0\}$ 称为状态 i 的占有时间过程。

现在我们假设 $X(0)=0$，并且要确定 $S_0(t)$ 的分布。从 (5.8.4) 算得它的均值：

$$E[S_0(t) \mid X(0) = 0] = \int_0^t E[1 - X(s)]ds$$

$$= \int_0^t P_{00}(s)ds$$

$$= \frac{\mu}{\lambda+\mu} t + \frac{\lambda}{(\lambda+\mu)^2}[1 - e^{-(\lambda+\mu)t}]$$

我们不想去计算 $S_0(t)$ 的方差（见习题 29），宁愿直接考虑其分布。

为了确定给定 $X(0) = 0$ 时 $S_0(t)$ 的条件分布，我们将利用 $\{X(t), t \geqslant 0\}$ 的一致化表示。开始先对 $N(t)$ 取条件得到，对 $s < t$，

$$P\{S_0(t) \leqslant s\} = \sum_{n=0}^{\infty} e^{-(\lambda+\mu)t} \frac{((\lambda+\mu)t)^n}{n!} P\{S_0(t) \leqslant s \mid N(t) = n\}$$

$$= \sum_{n=1}^{\infty} e^{-(\lambda+\mu)t} \frac{((\lambda+\mu)t)^n}{n!} P\{S_0(t) \leqslant s \mid N(t)=n\}$$

最后的等式是因为 $N(t)=0$ 蕴含着 $S_0(t)=t$(由于 $X(0)=0$)。在 $N(t)=n$ 的条件下,把区间 $(0,t)$ 分为 $n+1$ 个子区间 $(0,X_{(1)})$, $(X_{(1)},X_{(2)}),\cdots,(X_{(n-1)},X_{(n)}),(X_{(n)},t)$,其中 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ 是 $\{N(s)\}$ 到 t 时为止的 n 个事件的发生时刻。过程 $\{X(t)\}$ 在任一指定的子区间中处于同一状态。在第一个子区间中它处于状态 0,而在往后各个子区间中相互独立地以概率 $\mu/(\lambda+\mu)$ 处于状态 0。因此,给定 $N(t)=n$ 时,过程处于状态 0 的子区间的个数等于 1 加上一个贝努里 $(n,\mu/(\lambda+\mu))$ 随机变量。若此和等于 k(即,若有 $1+\text{Bin}(n,\mu/(\lambda+\mu))=k$) 则 $S_0(t)$ 将等于 k 个上述子区间长度之和。然而,因 $X_{(1)},\cdots,X_{(n)}$ 的分布如同 n 个独立的 $(0,t)$ 上均匀分布的随机变量的顺序统计量的分布,所以 $n+1$ 个子区间长 $Y_1,Y_2,\cdots,Y_{(n+1)}$ 的联合分布(其中 $Y_i = X_{(i)} - X_{(i-1)}, i=1,\cdots,n+1$,及 $X_{(0)}=0, X_{(n+1)}=t$) 是可交换的,也即 $P\{Y_i \leqslant y_i, i=1,\cdots,n+1\}$ 是向量 $(y_1,\cdots,y_{n+1})$ 的对称函数(见习题 30)。所以,Y_i 中任何 k 个之和的分布都相同,不管是哪些 Y_i 在和中。最后,因

$$X_{(k)} = Y_1 + \cdots + Y_k$$

可见当 $k \leqslant n$ 时,这和与 n 个独立 $(0,t)$ 上均匀分布的随机变量的第 k 个最大值有同样的分布。(当 $k=n+1$ 时此和等于 t) 因此,对 $s<t$

$$P\{X_{(k)} \leqslant s\} = \sum_{i=k}^{n} \binom{n}{i} \left(\frac{s}{t}\right)^i \left(1-\frac{s}{t}\right)^{n-i}$$

综上所述,对 $s<t$

$$P\{S_0(t) \leqslant s \mid X(0)=0\}$$

$$= \sum_{n=1}^{\infty} e^{-(\lambda+\mu)t} \frac{((\lambda+\mu)t)^n}{n!} \sum_{k=1}^{n} \binom{n}{k-1} \left(\frac{\mu}{\lambda+\mu}\right)^{k-1} \left(\frac{\lambda}{\lambda+\mu}\right)^{n-}$$

$$\times \sum_{i=k}^{n} \binom{n}{i} \left(\frac{s}{t}\right)^i \left(1-\frac{s}{t}\right)^{n-i}$$

也有

$$P\{S_0(t)=t\,|\,X(0)=0\}=e^{-\lambda t}$$

习 题

1. 一生物群体由雄性及雌性成员组成，在一个小的群落中任一特定的雄性成员与任一特定的雌性成员在任一长为 h 的时间区间中交配的概率为 $\lambda h+o(h)$。每次交配立即产生一个后代，它等可能地是雄性或雌性。令 $N_1(t)$ 与 $N_2(t)$ 表示在 t 时群体中雄性与雌性的个数。导出连续时间马尔可夫链 $\{N_1(t),N_2(t)\}$ 的参数。
2. 假设一个单细胞生物处于两个状态——A 或 B 之一，处于状态 A 的一个个体以指数率 α 变到状态 B；处于状态 B 的一个个体以指数率 β 分裂成两个新的 A 型个体。为这样的生物群体定义一个合适的连续时间马尔可夫链，并且决定这个模型的适当的参数。
3. 证明连续时间马尔可夫链是规则的，若已知 (a) 对一切 i 有 $\nu_i<M<\infty$ 或 (b) 转移概率为 P_{ij} 的离散时间马尔可夫链是不可约且常返的。
4. 对于一个具有出生参数 λ_n ($n\geqslant 0$) 的纯生过程，计算群体总数从 0 到 N 所需时间的均值、方差与矩母函数。
5. 考虑一个 $X(0)=i$ 的尤尔过程。给定 $X(t)=i+k$，关于在 $(0,t)$ 内出生的 k 个个体的出生时间的条件分布能说些什么？
6. 验证例 5.3 (b) 中给出的公式

$$A(t)=a_0+\int_0^t X(s)ds$$

7. 考虑一个从一个个体开始的尤尔过程，并假设在时刻 s 出生的个体以概率 $P(s)$ 是健壮的。计算在 $(0,t)$ 内出生的健壮的个体个数之分布。
8. 证明引理 5.4.1。
9. 证明引理 5.4.2。
10. 令 $P(t)=P_{00}(t)$

(a) 求

$$\lim_{t\to 0}\frac{1-P(t)}{t}$$

(b) 证明：

$$P(t)P(s) \leqslant P(t+s) \leqslant 1 - P(s) + P(s)P(t)$$

(c) 证明：

$$|P(t) - P(s)| \leqslant 1 - P(t-s), \qquad s < t$$

并得出 P 是连续的结论。

11. 对尤尔过程：

(a) 验证

$$P_{ij}(t) = \binom{j-1}{i-1} e^{-i\lambda t}(1 - e^{-\lambda t})^{j-i}$$

满足向前与向后方程。

(b) 假设 $X(0)=1$ 且在时刻 T 过程停止并由一个移出过程替代，其中移出按照速率为 μ 的泊松过程而发生。以 τ 记在 T 之后群体灭绝所需时间。求 τ 的密度函数并证明

$$E[\tau] = e^{\lambda T}/\mu$$

12. 假设系统的"状态"可建模为两状态的连续时间马尔可夫链，其转移率为 $\nu_0=\lambda$，$\nu_1=\mu$。当系统的状态是 i 时，"事件"按照速率为 α_i 的泊松过程发生，$i=0$，1。以 $N(t)$记$(0,t)$中事件的个数。

(a) 求$\lim_{t\to\infty} N(t)/t$。

(b) 如果初始状态是状态 0，求 $E[N(t)]$。

13. 假定一生物群体中的各个个体以指数率 λ 出生而以指数率 μ 死亡。另外，存在由迁入引起的指数增长率 θ，但是当群体总数是 N 或更大时，则不允许迁入。

(a) 对此建立一个生灭模型。

(b) 如果 $N=3$，$1=\theta=\lambda$，$\mu=2$，确定迁入受到限制的时间的比例。

14. 一家由一名理发员主持的小理发店，至多容纳两个人。可能来的顾客以每小时 3 人的速率的泊松过程来到，相继的服务时间是独立的均值为$\frac{1}{4}$小时的指数随机变量。

(a) 店中顾客的平均数是多少？

(b) 进店的顾客占多少比率？

(c) 若理发员能加快一倍地工作，他会多做多少生意？

15. 对 $M/M/s$ 系统求极限概率，并且确定极限概率存在所需的条件。

16. 若$\{X(t), t\geqslant 0\}$与$\{Y(t), t\geqslant 0\}$是独立的时间可逆的马尔可夫链，证明过

程$\{(X(t),Y(t)),t\geqslant 0\}$也是如此。

17. 考虑参数分别为λ_i，μ_i的两个$M/M/1$排队系统，其中$\lambda_i<\mu_i$，$i=1$，2。假设它们共同使用一个容纳N个人的候客室。(即每当候客室占满时一切来客都消失。)计算在第一个系统中有n人(当$n>0$时1人在接受服务而$n-1$人在候客室中)而m人在第二个系统的极限概率。(提示：利用习题16的结果。)

18. 关于只能容纳有限个人的平稳$M/M/1$系统的离去过程，你能说些什么？

19. 在5.6.2节的随机群体模型中：

(a) 证明：当$P(\underline{n})$由(5.6.4)给出，其中$\alpha_j=(\lambda/j\nu)(\nu/\mu)^j$时

$$P(\underline{n})q(\underline{n},D_j\underline{n}) = P(D_j\underline{n})q(D_j\underline{n},\underline{n})$$

(b) 以$D(t)$记$(0,t)$中灭绝的家族数。假设过程在时刻$t=0$处于稳定态0，$\{D(t),t\geqslant 0\}$是什么类型的随机过程？若最初群体在$t=0$是空的，它又是什么过程？

20. 完成5.7.1节的排队网络模型中的猜测的证明。

21. N个顾客游动于r个服务员之间。在i号服务员处的服务时间是参数为μ_i的指数变量。当一顾客离开i号服务员时他以概率$1/(r-1)$去j号服务员处($j\neq i$)排队(若有人在等待，否则他就直接进入服务)。令状态为$(n_1, \cdots, n_r)$，若在i号服务员处有n_i个顾客，$i=1$，…，r。证明相应的连续时间马尔可夫链是时间可逆的，并求极限概率。

22. 考虑一时间可逆的连续时间马尔可夫链，它有参数ν_i，P_{ij}及极限概率P_j，$j\geqslant 0$。选择某个状态(比如说状态0)，考虑一个新的马尔可夫链，它使状态0成为一个吸收状态，即重新令ν_0等于0。现在，假设在按照一速率为λ的泊松过程选取的时刻都有马尔可夫链开始，它们都是如上的类型(以0作为吸收状态)，以概率P_{0i}选取初始状态为j。

假定一切存在的链是独立的。以$N_j(t)$记时刻t处于状态j($j>0$)的链的个数。

(a) 论证如果没有初始链，则$N_j(t)(j>0)$是独立的泊松随机变量。

(b) 在稳态中论证向量过程$\{(N_1(t),N_2(t),\cdots)\}$是时间可逆的，具有平稳分布：对$\underline{n}=(n_1, n_2, \cdots)$

$$P(\underline{n}) = \prod_{j=1}^{\infty} e^{-\alpha_j}\frac{\alpha_j^{n_j}}{n_j!}$$

其中$\alpha_j=\lambda P_j/P_0\nu_0$。

23. 考虑一个 $M/M/\infty$ 系统，具有编号为 1，2，… 的通路（服务员），顾客一来到就挑选编号最小的空间通路。于是我们可以认为一切来到全发生于 1 号通路，发现 1 号通路忙着的顾客就溢出而变成 2 号通路的来到。发现 1 号与 2 号通路都忙着的顾客就溢出 2 号路，变成 3 号通路的来到，等等。

(a) 1 号通路的忙时占多少比率？

(b) 通过考虑相应的 $M/M/2$ 消失系统，确定 2 号通路忙时的比率。

(c) 对任意 c 写出 c 号通路忙时的比率的表达式。

(d) 从 c 号通路到 $c+1$ 号通路的溢出率是多少？相应的溢出过程是泊松过程吗？简要地解释。

(e) 若服务分布是一般的而不是指数的，你对(a) — (d)的回答中哪几个（如果有的话）要改动？简要地解释。

24. 证明定理 5.7.1。

25. (a) 证明平稳马尔可夫过程是可逆的，当且仅当它的转移率对任意有限状态序列 j_1，j_2，…，j_n 满足

$$q(j_1,j_2)q(j_2,j_3)\cdots q(j_{n-1},j_n)q(j_n,j_1)$$
$$= q(j_1,j_n)q(j_n,j_{n-1})\cdots q(j_3,j_2)q(j_2,j_1)$$

(b) 论证只要验证(a)中等式对相异状态的序列成立就够了。

(c) 假设来到一排队系统的顾客流形成一速率为 ν 的泊松过程且有两个工作效率可能不同的服务员。具体地，假设顾客在 i 号服务员处的服务时间服从参数为 μ_i 的指数分布，$i=1$，2，其中 $\mu_1+\mu_2>\nu$。若一来到的顾客发现两名服务员都有空，则他等可能地被分配到任一服务员那儿去。对此模型定义一个恰当的连续时间马尔可夫链，证明它是时间可逆的，且确定极限概率。

26. 一排队系统在任一时刻的工作量定义为该时刻系统中全体顾客的剩余服务时间之和。对稳态中的 $M/G/1$ 计算系统工作量的均值与方差。

27. 考虑一个处在稳态中的遍历连续时间马尔可夫链，其转移率为 q_{ij}。以 P_j $(j\geqslant 0)$ 记平稳分布。假设状态空间被分成两个子集 B 与 $B^c=G$。

(a) 计算过程在 B 中的条件下它处于状态 i $(i\in B)$的概率，即计算

$$P\{X(t)=i|X(t)\in B\}$$

(b) 计算过程刚进入 B 的条件下它处于状态 i $(i\in B)$ 的概率，即计算

$$P\{X(t)=i|X(t)\in B,X(t^-)\in G\}$$

(c) 对于 $i\in B$，以 T_i 记已知过程处于状态 i 直到它进入 G 为止的时间，

且令 $\widetilde{F}_i(s)=E\ [e^{-sT_i}]$。论证

$$\widetilde{F}_i(s)=\frac{\nu_i}{\nu_i+s}[\sum_{j\in B}\widetilde{F}_j(s)P_{ij}+\sum_{j\in G}P_{ij}]$$

其中 $P_{ij}=q_{ij}/\sum_{j\neq i}q_{ij}$。

(d) 论证

$$\sum_{i\in G}\sum_{j\in B}P_iq_{ij}=\sum_{i\in B}\sum_{j\in G}P_iq_{ij}$$

(e) 利用(c)与(d)证明：

$$s\sum_{i\in B}P_i\widetilde{F}_i(s)=\sum_{i\in G}\sum_{j\in B}P_iq_{ij}(1-\widetilde{F}_j(s))$$

(f) 设过程刚从 G 进入 B，以 T_ν 记直到它离开 B 为止的时间。利用 (b) 断定

$$E[e^{-sT_\nu}]=\frac{\sum_{i\in B}\sum_{j\in G}\widetilde{F}_i(s)P_jq_{ij}}{\sum_{j\in G}\sum_{k\in B}P_jq_{ik}}$$

(g) 利用(e)与(f)论证

$$\sum_{j\in B}P_j=[\sum_{i\in G}\sum_{j\in B}P_jq_{ij}]E[T_\nu]$$

(h) 设过程处于 B 中的一个状态，以 T_x 记直到它离开 B 为止的时间。利用 (a)，(e)，(f)及(g) 证明：

$$E[e^{-sT_x}]=\frac{1-E[e^{sT_\nu}]}{sE[T_\nu]}$$

(i) 利用 (h) 与拉普拉斯变换的唯一性断定

$$P\{T_x\leqslant t\}=\int_0^t\frac{P\{T_\nu>s\}ds}{E[T_\nu]}$$

(j) 利用(i)证明：

$$E[T_x]=\frac{E[T_\nu^2]}{2E[T_\nu]}\geqslant\frac{E[T_\nu]}{2}$$

随机变量 T_ν 称为在状态集 B 中的逗留时间，它表示到达 B 后在 B 中渡过的时间。随机变量 T_x，称为从 B 退出的时间，它表示已知过程现在是在 B 中但还要留在 B 中的剩余时间。上述结果表明，T_x 与 T_ν 的分布之间有着与稳态下的更新过程的剩余寿命的分布与相继的更新间的时间的分布之间的同样的结构关系。

28. 考虑一更新过程，其来到间隔分布 F 是两个指数分布的混合，即 $\widetilde{F}(x)$

$= pe^{-\lambda_1 x} + qe^{-\lambda_2 x}, q = 1 - p$。计算更新函数 $E[N(t)]$。

提示：设想在每个更新点掷一枚硬币，出正面的概率为 p。若出现正面，则下一来到间隔是参数为 λ_1 的指数变量；如果反面出现，则是参数为 λ_2 的指数变量。令 $R(t) = i$，若时刻 t 的指数率为 λ_i。则

(a) 确定 $P\{R(t) = i\}, \quad i = 1,2$

(b) 论证

$$E[N(t)] = \sum_{i=1}^{2} \lambda_i \int_0^t P\{R(s) = i\} ds = E[\int_0^t \Lambda(s) ds]$$

其中 $\Lambda(s) = \lambda_{R(s)}$。

29. 考虑例 5.8(a) 的两状态马尔可夫链，其 $X(0) = 0$。

(a) 计算 $\mathrm{Cov}(X(s), X(y))$。

(b) 以 $S_0(t)$ 记到 t 为止状态 0 的占有时间。利用(a) 与(5.8.4) 计算 $\mathrm{Var}(S(t))$。

30. 令 $Y_i = X_{(i)} - X_{(i-1)}, i = 1, \cdots, n+1$，其中 $X(0) = 0$，$X_{(n+1)} = t$，且 $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ 是 n 个独立的 $(0,t)$ 上均匀分布随机变量的顺序值。论证 $P\{Y_i \leqslant y_i, i = 1, \cdots, n+1\}$ 是 $y_1, \cdots, y_n$ 的对称函数。

参考文献

文献 1,2 及 3 提供了许多使用连续时间马尔可夫链的例子。关于时间可逆性的补充材料读者应参考文献 5 与 6。文献 6 始终一贯地在处理多种多样模型中利用逆向链的概念。它受到强烈的推荐。文献 7 提供了排队论的很好的导引与评述。

1. H. Bailey, *The Elements of Stochastic Processes with Application to the Natural Sciences*, Wiley, New York, 1964.
2. D. J. Bartholomew, *Stochastic Models for Social Processes*, 2nd ed., Wiley, London, 1973.
3. M. S. Bartlett, *An Introduction to Stochastic Processes*, 3rd ed., Cambridge University Press, Cambridge, England, 1978.
4. D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes*, Chapman and Hall, London, 1965.

5. J. Keilson, *Markov Chain Models—Rarity and Exponentiality*, Springer-Verlag, 1979.

6. F. Kelly, *Reversibility and Stochastic Networks*, Wiley, New York, 1979.

7. L. Kleinrock, *Queueing Systems*, Vols. I and II, Wiley, New York, 1975 and 1976.

6 布朗运动与其它的马尔可夫过程

6.1 引言及基本定义

我们从讨论对称随机游动开始，此游动每个单位时间等可能地向左或向右走一个单位步子。现在假设我们加速此过程，在越来越小的时间间隔中走越来越小的步子。若以正确的方式趋于极限，我们得到的就是布朗运动。

更精确地，假设每隔 Δt 时间等概率地向左或向右走一步，步子的大小取为 Δx。若以 $X(t)$ 记时刻 t 的位置，则

(6.1.1) $$X(t)=\Delta x(X_1+\cdots+X_{[t/\Delta t]})$$

其中

$$X_i=\begin{cases}+1, & \text{若长为 } \Delta x \text{ 的第 } i \text{ 步向右}\\ -1, & \text{若它向左}\end{cases}$$

且假设诸 X_i 相互独立，

$$P\{X_i=1\}=P\{X_i=-1\}=\frac{1}{2}$$

由于 $E[X_i]=0$，$\mathrm{Var}(X_i)=E[X_i^2]=1$，由(6.1.1)可见

(6.1.2) $$E[X(t)]=0$$

$$\mathrm{Var}(X(t))=(\Delta x)^2\left[\frac{t}{\Delta t}\right]$$

现在要令 Δx 与 Δt 趋于 0。然而，我们的做法必须使所得极限过程是非平凡的(例如，若令 $\Delta x=\Delta t$，再取 $\Delta t\to 0$，则从上面的讨论可见 $E[X(t)]$及 $\mathrm{Var}(X(t))$将同时收敛于 0，从而 $X(t)$要以概率 1 等于 0)。若令 $\Delta x=c\sqrt{\Delta t}$，$c$ 为某个正常数，则从(6.1.2)可见，$\Delta t\to 0$ 时

$$E[X(t)]=0$$

$$\mathrm{Var}(X(t))\to c^2t$$

现在我们列出取 $\Delta x=c\sqrt{\Delta t}$并令 $\Delta t\to 0$ 所得的极限过程的一些直观性质。从(6.1.1)及中心极限定理可见：

(i) $X(t)$是正态的，均值为 0，方差为 c^2t。

此外，由于随机游动的值在不相重叠的时间区间中的变化是独立的，我们有

(ii) $\{X(t),\ t\geqslant 0\}$有独立增量。

最后，由于随机游动在任一时间区间中的位置变化的分布只依赖于区间的长度，看来有

(iii) $\{X(t),\ t\geqslant 0\}$有平稳增量。

现在我们已为下列定义作好准备。

定义 随机过程$\{X(t),\ t\geqslant 0\}$ 称为布朗运动过程，若

(i) $X(0)=0$；

(ii) $\{X(t),\ t\geqslant 0\}$有平稳独立增量；

(iii) 对每个 $t>0$，$X(t)$ 服从正态分布，均值为 0，方差为 c^2t。

布朗运动过程，有时称为维纳过程，是应用概率论中最有用的随机过程之一。它起源自物理学中对布朗运动的一种描述，这种现象，以发现它的英国植物学家罗伯特·布朗命名，是一个完全浸没于一种液体或气体中的小粒子显示出的运动。自从它被发现以来，这过程被有效地应用于这样一些领域，如拟合优度的统计检验，分析股票市场的价格水平及量子力学等。

布朗运动现象的首次解释是爱因斯坦于 1905 年给出的。他证明，假设浸没的粒子连续不断地受到周围介质的分子的冲击，布

朗运动就可得到解释。然而上述简洁的用以描述布朗运动的随机过程的定义是维纳在起自 1918 年的一系列论文中给出的。

当$c=1$，这过程常被称为标准布朗运动。由于任一布朗运动总能转化为标准过程，为此只要看$X(t)/c$就行了，故今后将假定$c=1$。

把布朗运动解释为随机游动(6.1.1)的极限提示了$X(t)$应是t的连续函数。情况确是这样，可以证明以概率1，$X(t)$确实是t的连续函数，这个事实非常深刻，本书不给出其证明。我们也应指出，尽管得知$X(t)$的样本路径总是连续的，但它决不是普通的函数。因为正如从作为随机游动的极限的解释中可预料的，$X(t)$总是有尖角的，因此永远不会光滑，事实上能证明(虽然十分深奥)，以概率1，$X(t)$是无处可微的。

独立增量的假设蕴含了，在时刻s与$t+s$之间的位置变化(即$X(t+s)-X(s)$)与过程在时刻s之前的值独立。因此

$$\begin{aligned}
&P\{X(t+s)\leqslant a|X(s)=x,X(u),0\leqslant u<s\}\\
&=P\{X(t+s)-X(s)\leqslant a-x|X(s)=x,X(u),0\leqslant u<s\}\\
&=P\{X(t+s)-X(s)\leqslant a-x\}\\
&=P\{X(t+s)\leqslant a|X(s)=x\}
\end{aligned}$$

这表明给定现在$X(s)$及过去$X(u)(0<u<s)$将来的状态$X(t+s)$的条件分布只依赖于现在。满足这一条件的过程称为马尔可夫过程。

由于$X(t)$是正态的，均值为0，方差为t，它的密度函数为

$$f_t(x)=\frac{1}{\sqrt{2\pi t}}e^{-x^2/2t}$$

由平稳独立增量的假设容易得出，$X(t_1)$，…，$X(t_n)$的联合密度为

(6.1.3)

$$f(x_1,x_2,\cdots,x_n)=f_{t_1}(x_1)f_{t_2-t_1}(x_2-x_1)\cdots f_{t_n-t_{n-1}}(x_n-x_{n-1})$$

利用(6.1.3)，原则上我们可以计算任何想求的概率。例如，假

定我们要求给定 $X(t)=B$ 时 $X(s)$ 的条件分布，其中 $s<t$。条件密度是

$$f_{s/t}(x|B)=\frac{f_s(x)f_{t-s}(B-x)}{f_t(B)}$$
$$=K_1\exp\left\{\frac{-x^2}{2s}-\frac{(B-x)^2}{2(t-s)}\right\}$$
$$=K_2\exp\left\{-\frac{t(x-Bs/t)^2}{2s(t-s)}\right\}$$

因此给定 $X(t)=B$ 时 $X(s)(s<t)$ 的条件分布是正态的，其均值及方差为

(6.1.4a) $$E[X(s)|X(t)=B]=Bs/t$$

(6.1.4b) $$\mathrm{Var}(X(s)|X(t)=B)=s(t-s)/t$$

有意思的是注意到给定 $X(t)=B$ 时 $X(s)$ 的条件方差($s<t$)不依赖于 B! 也就是，若令 $s/t=\alpha$, $0<\alpha<1$，则给定 $X(t)$ 时 $X(s)$ 的条件分布是正态的，均值为 $\alpha X(t)$，方差为 $\alpha(1-\alpha)t$。

从(6.1.3)也得出，$X(t_1)$，$X(t_2)$，…，$X(t_n)$ 的联合分布是多元正态的，所以布朗运动过程是高斯过程，这里我们用到了下列定义。

定义 随机过程 $\{X(t),t\geqslant 0\}$ 称为高斯过程，若对一切 $t_1,\cdots,t_n$，$X(t_1),\cdots,X(t_n)$ 有多元正态分布。

由于多元正态分布完全由边际均值与协方差决定，所以布朗运动也能定义为一个高斯过程，其 $E[X(t)]=0$，对 $s\leqslant t$

$$\begin{aligned}\mathrm{Cov}(X(s),X(t))&=\mathrm{Cov}(X(s),X(s)+X(t)-X(s))\\&=\mathrm{Cov}(X(s),X(s))+\mathrm{Cov}(X(s),X(t)-X(s))\\&=s\end{aligned}$$

其中最后的等式来自独立增量性及 $\mathrm{Var}(X(s))=s$。

设 $\{X(t),t\geqslant 0\}$ 为布朗运动过程，考虑在 $X(1)=0$ 的条件下，过程在 0 与 1 之间的值，即考虑条件随机过程 $\{X(t),0\leqslant t\leqslant 1|X(1)=0\}$。利用与建立(6.1.4)相同的推导，我们能证明这过程是高斯过程，它以布朗桥著称(因为它同时在 0 与 1 被固定为 0)。我们

来计算它的协方差函数。因为由(6.1.4)，对 $s<1$，

$$E[X(s)\mid X(1)=0]=0$$

对 $s<t<1$ 我们有

$$\begin{aligned}
&\mathrm{Cov}\,[(X(s),X(t))\mid X(1)=0]\\
&\quad=E[X(s)X(t)\mid X(1)=0]\\
&\quad=E[E[X(s)X(t)\mid X(t),X(1)=0]\mid X(1)=0]\\
&\quad=E[X(t)E[X(s)\mid X(t)]\mid X(1)=0]\\
&\quad=E[X(t)\frac{s}{t}X(t)\mid X(1)=0]\quad(\text{由}(6.1.4a))\\
&\quad=\frac{s}{t}E[X^2(t)\mid X(1)=0]\\
&\quad=\frac{s}{t}t(1-t)\qquad(\text{由}(6.1.4b))\\
&\quad=s(1-t)
\end{aligned}$$

因此布朗桥可定义为均值为 0，协方差函数为 $s(1-t)(s\leqslant t)$ 的高斯过程，这导致得到这一过程的另一种方法。

命题 6.1.1

若 $\{X(t),t\geqslant 0\}$ 是布朗运动，则 $Z(t)=X(t)-tX(1)$ 时，$\{Z(t),0\leqslant t\leqslant 1\}$ 是布朗桥过程。

证明 由于 $\{Z(t),t\geqslant 0\}$ 显然是高斯过程，我们需要验证的只是 $E[Z(t)]=0$ 及 $s\leqslant t$ 时 $\mathrm{Cov}(Z(s),Z(t))=s(1-t)$。前者是显然的，后者得自

$$\begin{aligned}
\mathrm{Cov}(Z(s),Z(t))&=\mathrm{Cov}(X(s)-sX(1),X(t)-tX(1))\\
&=\mathrm{Cov}(X(s),X(t))-t\mathrm{Cov}(X(s),X(1))\\
&\quad-s\mathrm{Cov}(X(1),X(t))+st\mathrm{Cov}(X(1),X(1))\\
&=s-st-st+st\\
&=s(1-t)
\end{aligned}$$

证毕。

在研究经验分布函数时布朗桥起着关键作用。为明白这点，设 $X_1,X_2,\cdots$ 为独立的(0,1)上均匀分布的随机变量，并定义 $N_n(s)$ $(0<s<1)$ 为前几个变量中小于或等于 s 的个数，即

$$N_n(s) = \sum_{i=1}^{n} I_i(s)$$

其中

$$I_i(s) = \begin{cases} 1, & 若\ X_i \leqslant s \\ 0, & 其它 \end{cases}$$

随机函数 $F_n(s) = \frac{N_n(s)}{n}$ $(0 \leqslant s \leqslant 1)$ 称为经验分布函数。我们研究它在 $n \to \infty$ 时的极限性质。

因为 $N_n(s)$ 是参数为 n，s 的二项分布随机变量，由强大数定律得出，对固定的 s，$n \to \infty$ 时以概率 1

$$F_n(s) \to s$$

事实上，可以证明(所谓的格利汶科——康泰利定理)这收敛关于 s 是一致的。即以概率 1，$n \to \infty$ 时

$$\sup_{0<s<1} |F_n(s) - s| \to 0$$

由中心极限定理也得出，对固定的 s，$\sqrt{n}\,(F_n(s) - s)$ 为渐近正态分布，均值为 0，方差为 $s(1-s)$，即

$$P\{\alpha_n(s) < x\} \to \frac{1}{\sqrt{2\pi s(1-s)}} \int_{-\infty}^{x} \exp\left\{\frac{-y^2}{2s(1-s)}\right\} ds$$

其中

$$\alpha_n(s) = \sqrt{n}\,(F_n(s) - s)$$

我们来研究 $n \to \infty$ 时随机过程 $\{\alpha_n(s), 0 \leqslant s \leqslant 1\}$ 的极限性质。首先注意到，对 $s < t$，给定 $N_n(s)$ 时，$N_n(t) - N_n(s)$ 的条件分布正是参数为 $n - N_n(s)$ 及 $(t-s)/(1-s)$ 的二项分布。因此利用中心极限定理，$\alpha_n(s)$ 与 $\alpha_n(t)$ 的渐近联合分布看来应是二元正态分布，事实上，类似的推断可合理地预料到极限过程(若存在)是高斯过程。为了知道是哪一个高斯过程，我们需要计算 $E[\alpha_n(s)]$ 及 $\mathrm{Cov}(\alpha_n(s), \alpha_n(t))$。现在

$$E[\alpha_n(s)] = 0$$

对 $0 \leqslant s \leqslant t \leqslant 1$

$$\begin{aligned}\mathrm{Cov}\ (\alpha_n(s),\alpha_n(t))&=n\mathrm{Cov}(F_n(s),F_n(t))\\&=\frac{1}{n}\mathrm{Cov}(N_n(s),N_n(t))\\&=\frac{E[E[N_n(s)N_n(t)\,|\,N_n(s)]]-n^2st}{n}\\&=\frac{E\left[N_n(s)\left(N_n(s)+(n-N_n(s))\dfrac{t-s}{1-s}\right)\right]-n^2st}{n}\\&=s(1-t)\end{aligned}$$

利用 $N_n(s)$是参数为 n，s 的二项分布变量，再简化即得最后的等式。

因此极限随机过程是一个高斯过程，均值函数为 0，协方差函数为 $s(1-t)$，$0\leqslant s\leqslant t\leqslant 1$，看来是合理的（事实上能严格地证明）。但这正是布朗桥过程。

尽管上面的分析是在 X_i 有(0,1)上均匀分布的假设下进行的，它的适用范围可放宽，只要注意到若分布函数为 F，则 F 为连续时随机变量 $F(X_i)$在(0,1)上均匀分布。例如，假定我们要对任意的连续分布 F 研究

$$\sqrt{n}\sup_x|F_n(x)-F(x)|$$

的极限分布，其中 $F_n(x)$是分布为 F 的独立随机变量 X_i 的前 n 个中小于或等于 x 的比率。由前面的讨论，若令

$$\begin{aligned}\alpha_n(s)&=\sqrt{n}\,[(X_i,i=1,\cdots,n\ \text{中使}\ F(X_i)\leqslant s\ \text{的个数})-s]\\&=\sqrt{n}\,[(X_i,i=1,\cdots,n\ \text{中使}\ X_i\leqslant F^{-1}(s)\text{的个数})-s]\\&=\sqrt{n}\,[F_n(F^{-1}(s))-s]\\&=\sqrt{n}\,[F_n(y_s)-F(y_s)]\end{aligned}$$

其中 $y_s=F^{-1}(s)$，则 $\{\alpha_n(s),\ 0\leqslant s\leqslant 1\}$ 收敛于布朗桥过程。因此$\sqrt{n}\sup_x(F_n(x)-F(x))$的极限分布是布朗桥的上确界（或最大值，由于连续性）的分布。所以

$$\lim_{n\to\infty}P\{\sqrt{n}\sup_x|F_n(x)-F(x)|<a\}=P\{\max_{0\leqslant t\leqslant 1}|Z(t)|<a\}$$

其中$\{Z(t), 0\leqslant t\leqslant 1\}$为布朗桥过程。

6.2 击中时、最大值变量及反正弦律

以 T_a 记布朗运动过程首次击中 a 的时刻。当$a>0$ 时，为计算 $P\{T_a\leqslant t\}$，我们考虑 $P\{X(t)\geqslant a\}$并对 $T_a\leqslant t$ 是否发生取条件。这给出

(6.2.1)
$$P\{X(t)\geqslant a\}=P\{X(t)\geqslant a \mid T_a\leqslant t\}P\{T_a\leqslant t\} + P\{X(t)\geqslant a \mid T_a>t\}P\{T_a>t\}$$

若 $T_a\leqslant t$,则过程在$[0,t]$中的某个点击中 a,由对称性,在 t 时它在 a 之上或在 a 之下是等可能的,即

$$P\{X(t)\geqslant a \mid T_a\leqslant t\}=\frac{1}{2}$$

由于(6.2.1)中右边的第二项显然等于 0(因为由连续性,过程的值不可能还未击中 a 就大于 a),可见

(6.2.2)
$$\begin{aligned}P\{T_a\leqslant t\}&=2P\{X(t)\geqslant a\}\\&=\frac{2}{\sqrt{2\pi t}}\int_a^\infty e^{-x^2/2t}dx\\&=\frac{2}{\sqrt{2\pi}}\int_{a/\sqrt{t}}^\infty e^{-y^2/2}dy, \qquad a>0\end{aligned}$$

由此可见

$$P\{T_a<\infty\}=\lim_{t\to\infty}P\{T_a\leqslant t\}=\frac{2}{\sqrt{2\pi}}\int_0^\infty e^{-y^2/2}dy=1$$

此外,利用(6.2.2)我们得到

$$\begin{aligned}E[T_a]&=\int_0^\infty P\{T_a>t\}dt\\&=\int_0^\infty\left(1-\frac{2}{\sqrt{2\pi}}\int_{a/\sqrt{t}}^\infty e^{-y^2/2}dy\right)dt\\&=\frac{2}{\sqrt{2\pi}}\int_0^\infty\int_0^{a/\sqrt{t}} e^{-y^2/2}\,dydt\end{aligned}$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^\infty \int_0^{a^2/y^2} dt\, e^{-y^2/2}\, dy$$
$$= \frac{2a^2}{\sqrt{2\pi}} \int_0^\infty \frac{1}{y^2}\, e^{-y^2/2}\, dy$$
$$\geqslant \frac{2a^2 e^{-1/2}}{\sqrt{2\pi}} \int_0^1 \frac{1}{y^2}\, dy$$
$$= \infty$$

由此得出，T_a 虽然以概率 1 为有限，但有无穷的期望。也就是布朗运动过程以概率 1 迟早击中 a，但它的平均时间是无穷的。(这直观吗？想想对称随机游动。)

对 $a<0$，由对称性，T_a 的分布与 T_{-a}的分布相同。因此由(6.2.2)我们得到

$$P\{T_a \leqslant t\} = \frac{2}{\sqrt{2\pi}} \int_{|a|/\sqrt{t}}^\infty e^{-y^2/2}\, dy$$

另一个感兴趣的随机变量是过程在$[0,t]$中达到的最大值。它的分布可如下得到。对 $a>0$

$$P\{\max_{0\leqslant s\leqslant t} X(s) \geqslant a\} = P\{T_a \leqslant t\} \quad (\text{由连续性})$$
$$= 2P\{X(t) \geqslant a\}$$
$$= \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^\infty e^{-y^2/2}\, dy$$

以 $0(t_1,t_2)$记布朗运动过程在区间(t_1,t_2)中至少取到 0 一次这一事件。为计算 $P\{0(t_1,t_2)\}$，我们对 $X(t_1)$取条件如下：

$$P\{0(t_1,t_2)\} = \frac{1}{\sqrt{2\pi t_1}} \int_{-\infty}^\infty P\{0(t_1,t_2) \mid X(t_1)=x\} e^{-x^2/2t_1}\, dx$$

利用布朗运动关于原点的对称性及路径的连续性得

$$P\{0(t_1,t_2) \mid X(t_1)=x\} = P\{T_{|x|} \leqslant t_2-t_1\}$$

因此，利用(6.2.2)我们得到

$$P\{0(t_1,t_2)\} = \frac{1}{\pi\sqrt{t_1(t_2-t_1)}} \int_0^\infty \int_x^\infty e^{-y^2/2(t_2-t_1)}\, dy\, e^{-x^2/2t_1}\, dx$$

上列积分能算得明显表达式

$$P\{0(t_1,t_2)\}=1-\frac{2}{\pi}\arcsin\sqrt{t_1/t_2}$$

所以我们证明了下列

命题 6.2.1

对 $0<x<1$

$$P\{在(xt,t)中没有零点\}=\frac{2}{\pi}\arcsin\sqrt{x}$$

注记　命题 6.2.1 并不使我们惊讶。因为由第三章 3.7 节末尾的注记我们知道，对于对称随机游动

$$P\{在(nx,n)中设有零点\}\approx\frac{2}{\pi}\arcsin\sqrt{x}$$

当 $n\to\infty$ 时近似式成为精确等式。由于布朗运动是对称随机游动跳跃越来越快（跃度越来越小）的极限情形，直观上看来，对布朗运动上述近似式应成为等式。命题 6.2.1 证实了这一点。

也能证明第三章 3.7 节的另一个反正弦律（对称随机游动为正的时间的比率取极限时服从反正弦律）对布朗运动精确地成立，即能证明下列命题

命题 6.2.2

对布朗运动，以 $A(t)$ 记在 $[0,t]$ 中过程为正的时间，则对 $0<x<1$，

$$P\{A(t)/t\leqslant x\}=\frac{2}{\pi}\arcsin\sqrt{x}$$

6.3 布朗运动的各种变化

这一节中我们讨论布朗运动的四种变化，第一种假设布朗运动一达到某个给定的值就被吸收，第二种假定它不能变为负的，第三种处理几何布朗运动，而第四种为积分型的。

6.3.1 在一个值处被吸收的布朗运动

设$\{X(t)\}$为布朗运动，T_x 为它首次击中 x 的时刻，$x>0$。记 $Z(t)$为

$$Z(t)=\begin{cases}X(t), & 若\ t<T_x \\ x, & 若\ t\geqslant T_x\end{cases}$$

则$\{Z(t),t\geqslant 0\}$是击中 x 后永远停留在那里的布朗运动。

随机变量 $Z(t)$的分布既有离散又有连续部分。离散部分是

$$P\{Z(t)=x\}=P\{T_x\leqslant t\}$$

$$=\frac{2}{\sqrt{2\pi t}}\int_x^{\infty}e^{-y^2/2t}\,dy \qquad (由(6.2.2))$$

对于连续部分，对 $y<x$ 我们有

(6.3.1) $$P\{Z(t)\leqslant y\}=P\{X(t)\leqslant y,\max_{0\leqslant s\leqslant t}X(s)<x\}$$

$$=P\{X(t)\leqslant y\}-P\{X(t)\leqslant y,\max_{0\leqslant s\leqslant t}X(s)>x\}$$

计算右边第二项如下：

(6.3.2) $$P\{X(t)\leqslant y,\max_{0\leqslant s\leqslant t}X(s)>x\}$$

$$=P\{X(t)\leqslant y\,|\max_{0\leqslant s\leqslant t}X(s)>x\}P\{\max_{0\leqslant s\leqslant t}X(s)>x\}$$

现在注意到事件$\max_{0\leqslant s\leqslant t}X(s)>x$ 等价于事件 $T_x<t$；若布朗运动过程在时刻 $T_x(T_x<t)$击中 x，那么为了在时刻 t 小于 y，它在之后的 $t-T_x$ 时间中至少必须减少 $x-y$。由对称性，它与增加这个数量的可能性是相同的。因此

(6.3.3) $$P\{X(t)\leqslant y\,|\max_{0\leqslant s\leqslant t}X(s)>x\}$$

$$=P\{X(t)\geqslant 2x-y\,|\max_{0\leqslant s\leqslant t}X(s)>x\}$$

从(6.3.2)及(6.3.3)我们有

$$P\{X(t)\leqslant y,\max_{0\leqslant s\leqslant t}X(s)>x\}$$

$$=P\{X(t)\geqslant 2x-y,\max_{0\leqslant s\leqslant t}X(s)>x\}$$

$$=P\{X(t)\geqslant 2x-y\} \qquad (因为\ y<x)$$

由(6.3.1)

$$P\{Z(t)\leqslant y\}=P\{X(t)\leqslant y\}-P\{X(t)\geqslant 2x-y\}$$

$$=P\{X(t)\leqslant y\}-P\{X(t)\leqslant y-2x\} \qquad \text{（由正态分布的对称性）}$$

$$=\frac{1}{\sqrt{2\pi t}}\int_{y-2x}^{y} e^{-u^2/2t}\,du$$

6.3.2 在原点反射的布朗运动

若$\{X(t),t\geqslant 0\}$是布朗运动，则过程$\{Z(t),t\geqslant 0\}$，其中

$$Z(t)=|X(t)|, \qquad t\geqslant 0$$

称为在原点反射的布朗运动。

$Z(t)$的分布容易得到。对 $y>0$

$$P\{Z(t)\leqslant y\}=P\{X(t)\leqslant y\}-P\{X(t)\leqslant -y\}$$
$$=2P\{X(t)\leqslant y\}-1$$
$$=\frac{2}{\sqrt{2\pi t}}\int_{-\infty}^{y} e^{-x^2/2t}\,dx-1$$

其中最后第二个等式成立是因为 $X(t)$为零均值的正态变量。

容易计算 $Z(t)$的均值及方差且

(6.3.4) $$E[Z(t)]=\sqrt{2t/\pi}, \qquad \mathrm{Var}(Z(t))=\left(1-\frac{2}{\pi}\right)t$$

6.3.3 几何布朗运动

若$\{X(t),t\geqslant 0\}$是布朗运动，则由

$$Y(t)=e^{X(t)}$$

定义的过程$\{Y(t),t\geqslant 0\}$称为几何布朗运动。

由于 $X(t)$是均值为 0、方差为 t 的正态变量，它的矩母函数为

$$E[e^{sX(t)}]=e^{ts^2/2}$$

所以

$$E[Y(t)]=E[e^{X(t)}]=e^{t/2},$$
$$\mathrm{Var}(Y(t))=E[Y^2(t)]-(E[Y(t)])^2$$
$$=E[e^{2X(t)}]-e^{t}$$
$$=e^{2t}-e^{t}$$

在建模中认为百分比变化（不是绝对变化）是独立同分布时几何布朗运动是有用的。例如，假设 $Y(n)$是某种商品在时刻 n 的价

格。假设 $Y(n)/Y(n-1)$（而不是 $Y(n)-Y(n-1)$）独立同分布可能是合理的。令

$$X_n = Y(n)/Y(n-1)$$

取 $Y(0)=1$，则

$$Y(n) = X_1 X_2 \cdots X_n$$

所以

$$\log Y(n) = \sum_{i=1}^{n} \log X_i$$

由于 X_i 是独立同分布的，$\log Y(n)$ 经适当正则化后近似于布朗运动，从而 $\{Y(n)\}$ 近似于几何布朗运动。

例 6.3(a)　股票期权的价值。设一个人拥有在将来的一个时刻 T 以固定的价格 K 购买一股某种股票的期权。假设股票目前的价格为 y，且它的价格按照几何布朗运动变化，我们来计算拥有这期权的平均价值。若时刻 T 的股票价格是 K 或更高时，期权将被实施，因此它的平均价值为

$$\begin{aligned} E[\max(Y(T)-K,0)] &= \int_0^{\infty} P\{Y(T)-K>a\}da \\ &= \int_0^{\infty} P\{ye^{X(T)}-K>a\}da \\ &= \int_0^{\infty} P\left\{X(T)>\log\frac{K+a}{y}\right\}da \\ &= \frac{1}{\sqrt{2\pi T}} \int_0^{\infty}\int_{\log[(k+a)/y]}^{\infty} e^{-x^2/2T}dx\, da \end{aligned}$$

6.3.4　积分布朗运动

若 $\{X(t),t\geqslant 0\}$ 是布朗运动，则由

(6.3.5)
$$Z(t) = \int_0^t X(s)ds$$

定义的过程 $\{Z(t),t\geqslant 0\}$ 称为积分布朗运动。为了说明这样的过程如何在实际中发生，我们有意建立一种商品的价格随时间变化的模型，以 $Z(t)$ 记在时刻 t 的价格，与其假设 $\{Z(t)\}$ 是布朗运动（或几何布朗运动），我们宁愿假设 $Z(t)$ 的变化率遵循一个布朗运动。例如，我们可以假设商品价格的变化率是通货膨胀率，它可设想为如同布朗运动那样变化。因此

$$\frac{d}{dt}Z(t)=X(t)$$

或

$$Z(t)=Z(0)+\int_0^t X(s)ds$$

从布朗运动是高斯过程可得$\{Z(t),t\geqslant 0\}$也是高斯过程。为证此事首先回忆一下，W_1，…，W_n称为有联合正态分布，若它们可表示为

$$W_i=\sum_{j=1}^m a_{ij}U_j, \qquad i=1,\cdots,n$$

其中U_j（$j=1$，…，m）是独立正态随机变量。由此可得W_1，…，W_n的任意一组部分和也是联合正态的。把(6.3.5)中的积分写成近似和的极限就能证明$Z(t_1),\cdots,Z(t_n)$联合正态。

由于$\{Z(t),t\geqslant 0\}$是高斯过程，它的分布由其均值及协方差函数刻划。现在我们就计算它们：

$$\begin{aligned}E[Z(t)] &= E\left[\int_0^t X(s)ds\right]\\ &= \int_0^t E[X(s)]ds\\ &=0\end{aligned}$$

对$s\leqslant t$

(6.3.6)
$$\begin{aligned}\mathrm{Cov}[Z(s),Z(t)]&=E[Z(s)Z(t)]\\ &=E\left[\int_0^s X(y)dy\int_0^t X(u)du\right]\\ &=E\left[\int_0^s\int_0^t X(y)X(u)dydu\right]\\ &=\int_0^s\int_0^t E[X(y)X(u)]dydu\\ &=\int_0^s\int_0^t \min(y,u)dydu\\ &=\int_0^s\left(\int_0^u ydy+\int_u^t udy\right)du\\ &=s^2\left(\frac{t}{2}-\frac{s}{6}\right)\end{aligned}$$

由(6.3.5)定义的$\{Z(t), t\geqslant 0\}$不是马尔可夫过程。(为什么不是?) 然而，向量过程$\{(Z(t), X(t)) t\geqslant 0\}$是马尔可夫过程。(为什么?) 我们能计算$Z(t), X(t)$的联合分布，只要先注意到根据与前面相同的推理可知，它们是联合正态的。为了计算它们的协方差我们利用(6.3.6)如下：

$$\begin{aligned}
&\mathrm{Cov}\,(Z(t), Z(t)-Z(t-h)) \\
&\quad =\mathrm{Cov}(Z(t), Z(t))-\mathrm{Cov}(Z(t), Z(t-h)) \\
&\quad =\frac{t^3}{3}-(t-h)^2\left[\frac{t}{2}-\frac{t-h}{6}\right] \\
&\quad =t^2h/2+o(h)
\end{aligned}$$

然而

$$\begin{aligned}
&\mathrm{Cov}\,(Z(t), Z(t)-Z(t-h)) \\
&\quad =\mathrm{Cov}\left(Z(t), \int_{t-h}^{t} X(s)ds\right) \\
&\quad =\mathrm{Cov}(Z(t), hX(t)+o(h)) \\
&\quad =h\mathrm{Cov}(Z(t), X(t))+o(h)
\end{aligned}$$

所以

$$\mathrm{Cov}(Z(t), X(t))=t^2/2$$

因此，$Z(t), X(t)$有二元正态分布，且

$$E[Z(t)]=E[X(t)]=0$$

$$\mathrm{Cov}(X(t), Z(t))=t^2/2$$

若我们假设价格的百分比变化率遵循一个布朗运动过程，可得另一种形式的积分布朗运动。也就是，若以$W(t)$记t时的价格，则

$$\frac{d}{dt}W(t)=X(t)W(t)$$

或

$$W(t)=W(0)\exp\left\{\int_0^t X(s)ds\right\}$$

其中$\{X(t)\}$是布朗运动。取$W(0)=1$，我们看到

$$W(t)=e^{Z(t)}$$

由于 $Z(t)$ 是正态的，均值为 0，方差为 $t^2\left(\frac{t}{2}-\frac{t}{6}\right)=t^3/3$，可见

$$E[W(t)]=\exp\{t^6/6\}$$

6.4 有漂移的布朗运动

我们说 $\{X(t),t\geqslant 0\}$ 是漂移系数为 μ 的布朗运动过程，若

(i) $X(0)=0$；

(ii) $\{X(t),t\geqslant 0\}$ 有平稳独立增量；

(iii) $X(t)$ 服从均值为 μt，方差为 t 的正态分布。

我们也能定义它为 $X(t)=B(t)+\mu t$，其中 $\{B(t)\}$ 是标准布朗运动。

因此有漂移的布朗运动是一个以速率 μ 漂移开去的过程。如同布朗运动那样，它也能定义为随机游动的极限。为了明瞭这一点，假设每隔 Δt 时间，过程分别以概率 p 及 $1-p$ 按正方向或负方向走长为 Δx 的一步。若令

$$X_i=\begin{cases}1, & \text{若第 } i \text{ 步是正方向的}\\ -1, & \text{其它}\end{cases}$$

则时刻 t 的位置 $X(t)$ 是

$$X(t)=\Delta x(X_1+\cdots+X_{[t/\Delta t]})$$

现在

$$E[X(t)]=\Delta x[t/\Delta t](2p-1)$$

$$\mathrm{Var}(X(t))=(\Delta x)^2[t/\Delta t][1-(2p-1)^2]$$

因此，若令 $\Delta x=\sqrt{\Delta t}$，$p=\frac{1}{2}(1+\mu\sqrt{\Delta t})$，$\Delta t\to 0$，则

$$E[X(t)]\to\mu t$$

$$\mathrm{Var}[X(t)]\to t$$

但事实上 $\{X(t)\}$ 收敛于漂移系数为 μ 的布朗运动。

现在我们对此过程计算一些感兴趣的量，我们从过程在击中 $-B$ 之前击中 A 的概率开始，A，$B>0$。以 $P(x)$ 记这事件在过程

现在处于 x，$-B<x<A$ 的条件下的条件概率，即

$$P(x)=P\{X(t)\text{在击中}-B\text{之前击中}A|X(0)=x\}$$

对过程在时刻 0 与 h 之间的变化 $Y=X(h)-X(0)$ 取条件我们将得到一个微分方程。这给出

$$P(x)=E[P(x+Y)]+o(h)$$

上面的 $o(h)$ 是到时刻 h 过程早已击中 A 或 $-B$ 中的一个的概率。假设 $P(y)$ 在 x 点附近有泰勒级数展开，形式地进行可得

$$P(x)=E[P(x)+P'(x)Y+P''(x)Y^2/2+\cdots]+o(h)$$

由于 Y 是正态的，均值为 μh，方差为 h，我们得到

(6.4.1)

$$P(x)=P(x)+P'(x)\mu h+P''(x)\frac{\mu^2h^2+h}{2}+o(h)$$

因为所有大于二阶的微分项之和的均值是 $o(h)$。由(6.4.1)有

$$P'(x)\mu+\frac{P''(x)}{2}=\frac{o(h)}{h}$$

令 $h\to 0$

$$P'(x)\mu+\frac{P''(x)}{2}=0$$

将上式积分，我们得到

$$2\mu P(x)+P'(x)=c_1$$

或等价地

$$e^{2\mu x}(2\mu P(x)+P'(x))=c_1e^{2\mu x}$$

或

$$\frac{d}{dx}(e^{2\mu x}P(x))=c_1e^{2\mu x}$$

一积分即得

$$e^{2\mu x}P(x)=C_1e^{2\mu x}+C_2$$

因此

$$P(x)=C_1+C_2e^{-2\mu x}$$

利用边界条件 $P(A)=1$，$P(-B)=0$，能解得 C_1 及 C_2

$$C_1=\frac{e^{2\mu B}}{e^{2\mu B}-e^{-2\mu A}},\quad C_2=\frac{-1}{e^{2\mu B}-e^{-2\mu A}}$$

从而

(6.4.2) $$P(x)=\frac{e^{2\mu B}-e^{-2\mu x}}{e^{2\mu B}-e^{-2\mu A}}$$

因此从 $x=0$ 出发，在达到 $-B$ 之前先达到 A 的概率 $P(0)$ 为

(6.4.3) $$P\{\text{过程在下降 } B \text{ 之前先上升 } A\}=\frac{e^{2\mu B}-1}{e^{2\mu B}-e^{-2\mu A}}$$

注记

(1)等式(6.4.3)也能用对随机游动取极限的方法得到。因为由赌徒输光问题(见第四章的例 4.4(a))得出，当每局赌博分别以概率 p 及 $1-p$ 上升或下降 Δx 单位时，在下降 B 之前上升 A 的概率为

(6.4.4) $$P\{\text{下降 } B \text{ 之前上升 } A\}=\frac{1-\left(\frac{1-p}{p}\right)^{B/\Delta x}}{1-\left(\frac{1-p}{p}\right)^{(A+B)/\Delta x}}$$

当 $p=\frac{1}{2}(1+\mu\Delta x)$ 时，我们有

$$\lim_{\Delta x\to 0}\left(\frac{1-p}{p}\right)^{1/\Delta x}=\lim_{\Delta x\to 0}\left(\frac{1-\mu\Delta x}{1+\mu\Delta x}\right)^{1/\Delta x}$$
$$=e^{-\mu}/e^{\mu}$$

因此，令 $\Delta x\to 0$，由(6.4.4)可见

$$P\{\text{下降 } B \text{ 之前上升 } A\}=\frac{1-e^{-2\mu B}}{1-e^{-2\mu(A+B)}}$$

它与(6.4.3)相符。

(2)若 $\mu<0$，由(6.4.3)，令 B 趋于无穷可见

(6.4.5) $$P\{\text{过程迟早上升 } A\}=e^{2\mu A}$$

因此在这情形，过程漂向负无穷，而它的最大值是参数为 -2μ 的指数变量。

例 6.4(a) 实施股票期权。 假设我们有在将来的某个时刻以固定的价格 A 购买一股股票的期权，与现在的市价无关。股票现在的市价取为 0，并假

设它的变化遵循有负漂移系数$-d(d>0)$的布朗运动过程。问题是什么时候(若迟早要做)实施期权?

考虑在市价为x时实施期权的策略。在此策略下我们的平均所得是

$$P(x)(x-A)$$

其中$P(x)$是过程迟早到达x的概率。由(6.4.5)可见

$$P(x)=e^{-2dx}, \qquad x>0$$

x的最优值是使$(x-A)e^{-2dx}$最大的值,易见它是

$$x=A+1/2d$$

在等式(6.4.3)中令$\mu\to 0$可得,对布朗运动

(6.4.6) $$P\{\text{布朗运动在下降 } B \text{ 之前上升 } A\}=\frac{B}{A+B}$$

例 6.4 (b)　最优加倍策略。考虑两个人为一笔赌注玩某种随机游戏,最终一个选手成为胜者。起先一个选手被指定为"加倍者",这意味着他在任何时刻有权把赌注加倍。若在一个时刻他行使这个权利,那么另一个选手或者退出游戏,并将现在的赌注付给加倍者,或者同意将赌注加倍并继续游戏。若另一个选手决定继续玩下去,他就成为"加倍者"。换句话说,每一次加倍者行使其权利,加倍权就转给另一个选手。

我们假设游戏是观察从$\frac{1}{2}$出发的布朗运动。若它在到达0之前先击中1,则选手Ⅰ获胜,反之则选手Ⅱ获胜。由(6.4.6)知,若现在的状态是x,游戏继续进行到结束,则选手Ⅰ以概率x获胜。每一个选手的目标是使其平均获益最大,且我们假定每一个选手在博弈论意义下最优地玩游戏(例如,这意味着一个选手能宣布他的策略,而另一个选手即使知道这个信息也无法玩得更好些)。

直观上很明显最优策略应是下述类型的。

最优策略　假设选手Ⅰ(Ⅱ)有加倍权,则选手Ⅰ(Ⅱ)在时刻t加倍当且仅当$X(t)\geq p^*(X(t)\leq 1-p^*)$。选手Ⅱ(Ⅰ)的最优策略是在时刻$t$接受加倍当且仅当$X(t)\leq p^{**}(X(t)\geq 1-p^{**})$。剩下的是计算$p^*$及$p^{**}$。

引理 1　$p^*\leq p^{**}$

证明　对任意的$p>p^{**}$,若$X(t)=p$,选手Ⅰ加倍,则选手Ⅱ就退出。因此在$X(t)=p$时选手Ⅰ行使加倍权就能保证现有的赌注成为他的平均收益,而由于选手Ⅱ总能保证选手Ⅰ不可能获得更多(若选手Ⅰ加倍就退出),故得出选手Ⅰ行使加倍权必定是最优的。因此$p^*\leq p^{**}$。

引理 2　$p^{*}=p^{**}$.

证明　假定 $p^{*}<p^{**}$。我们要证明选手Ⅰ有比 p^{*} 更好的策略，从而得出矛盾。具体地说，比起在 p^{*} 加倍，选手Ⅰ更好的做法宁愿是等待 $X(t)$ 击中 0 或 p^{**}。如果它击中 p^{**}，那么选手Ⅰ能加倍，而由于选手Ⅱ将接受，所以选手Ⅰ的状况与在 p^{*} 加倍完全一样。另一方面，若在到达 p^{**} 之前击中 0，那么在新策略之下，他只输掉原来的赌注，而在 p^{*} 策略之下，他要输掉加倍的赌注。

因此由引理 1 及 2，只有一个临界值 p^{*}，使得若选手Ⅰ有加倍权，则他的最优策略是在时刻 t 加倍当且仅当 $X(t)\geqslant p^{*}$。类似地，选手Ⅱ的最优策略是在时刻 t 接受当且仅当 $X(t)\leqslant p^{*}$。由连续性可得在状态为 p^{*} 时，两个选手对他们的选择都是不在乎的。我们将利用这一点来计算 p^{*}。

设赌注是一个单位。现在若选手Ⅰ在 p^{*} 加倍，那么选手Ⅱ对退出还是接受加倍并不在乎都一样。由于在前一种选择下选手Ⅰ赢得 1，我们有

$$1=E[\text{若选手Ⅱ在 } p^{*} \text{ 接受，选手Ⅰ的收益}]$$

现在若选手Ⅱ在 p^{*} 接受，那么Ⅱ有下一次加倍权，只要 $X(t)$ 击中 $1-p^{*}$ 他将行使这特权。如果永远没有击中 $1-p^{*}$（即先击中 1），则Ⅱ输掉 2 个单位。由于从 p^{*} 出发在到达 1 之前击中 $1-p^{*}$ 的概率按(6.4.6)是 $(1-p^{*})/p^{*}$，我们有

$$1=E[\text{Ⅰ 的收益}\mid\text{击中 }1-p^{*}]\frac{1-p^{*}}{p^{*}}+2\frac{2p^{*}-1}{p^{*}}$$

如果击中 $1-p^{*}$，那么Ⅱ将加倍赌注到 4 个单位，而Ⅰ对接受与否并不在乎都一样。因此若Ⅰ退出，则他输掉 2，我们有

$$E[\text{Ⅰ 的收益}\mid\text{击中 }1-p^{*}]=-2$$

所以

$$1=-2\frac{1-p^{*}}{p^{*}}+2\frac{2p^{*}-1}{p^{*}}$$

或

$$p^{*}=\frac{4}{5}$$

例 6.4 (c)。　控制生产过程。本例中我们考虑一个随时间而趋于恶化的生产过程。具体地说，我们假设生产过程状态的变化遵循一个漂移系数为 μ，$\mu>0$，的维纳过程。当过程的状态为 B 时，假定过程就停顿，必须付出

费用 R 使过程回复到状态 0。另一方面，在达到停顿点 B 之前，我们可以试图修理。若状态为 x 且作了尝试去修理，这个尝试以概率 α_x 成功，以概率 $1-\alpha_x$ 失败。若尝试成功，过程回到状态 0，若尝试失败，假定过程到达 B（即过程停顿）。尝试一次修理的费用是 C。

我们要决定使得长时间之后的每单位时间的平均费用为最小的策略，且这么做时，我们将限于过程状态为 x，$0<x<B$，时尝试修理的策略。对这些策略，回到状态 0 显然构成一次更新，因此由第三章的定理 3.6.1 平均费用正是

(6.4.7) $$\frac{E[\text{一个循环的费用}]}{E[\text{一个循环的长度}]}=\frac{C+R(1-\alpha_x)}{E[\text{到达 } x \text{ 的时间}]}$$

以 $f(x)$ 记使过程到达 x 所需的平均时间。对时间 h 中的变化 $Y=X(h)-X(0)$ 取条件，我们可推导出一个 $f(x)$ 的微分方程。这给出

$$f(x)=h+E[f(x-Y)]+o(h)$$

其中 $o(h)$ 项表示到时间 h 过程早已击中 x 的概率。展开成泰勒级数

$$\begin{aligned}f(x)&=h+E\left[f(x)-Yf'(x)+\frac{Y^2}{2}f''(x)+\cdots\right]+o(h)\\&=h+f(x)-\mu hf'(x)+\frac{h}{2}f''(x)+o(h)\end{aligned}$$

或等价地

$$1=\mu f'(x)-\frac{f''(x)}{2}+\frac{o(h)}{h}$$

令 $h\to 0$

(6.4.8) $$1=\mu f'(x)-f''(x)/2$$

与其直接去解上述方程，不如先注意到

$$\begin{aligned}f(x+y)&=E[\text{从 0 到 } x+y \text{ 的时间}]\\&=E[\text{到 } x \text{ 的时间}]+E[\text{从 } x \text{ 到 } x+y \text{ 的时间}]\\&=f(x)+f(y)\end{aligned}$$

从而 $f(x)$ 具有形式 $f(x)=cx$，并由(6.4.8)可见 $c=1/\mu$。所以

$$f(x)=x/\mu$$

因此由(6.4.7)，状态为 x，$0<x<B$，时尝试修理的策略的长时间之后的平均费用为

$$\frac{\mu[C+R(1-\alpha_x)]}{x}$$

而从不试图修理的策略的长时间之后的平均费用为

$$R\mu/B$$

对于一个给定的函数 α_x，我们能用微积分的方法去决定使长时间之后的平均费用最小的策略。

以 T_x 记使漂移系数为 μ 的布朗运动过程击中 x 的时间(若过程永不击中 x，令它为∞)。我们将对 $x>0$ 计算它的矩母函数 $E[e^{-\theta}T_x]$，$\theta>0$，与在例6.4 (c) 中计算 $E\,[T_x]$ 的方法大体相同。先注意

(6.4.9)

$$\begin{aligned}E[\exp\{-\theta T_{x+y}\}]&=E[\exp\{-\theta(T_x+T_{x+y}-T_x)\}]\\&=E[\exp\{-\theta T_x\}]E[\exp\{-\theta(T_{x+y}-T_x)\}]\\&=E[\exp\{-\theta T_x\}]E[\exp\{-\theta T_y\}]\end{aligned}$$

其中最后一个等式得自平稳性，最后第二个等式得自独立增量性。但(6.4.9)蕴含了对某个 $c>0$

$$E[e^{-\theta Tx}] = e^{-cx}$$

为了决定 c，令

$$f(x) = E[e^{-\theta Tx}]$$

对 $Y = X(h) - X(0)$ 取条件将得到 f 满足的微分方程。这样做给出

$$\begin{aligned}f(x) &= E[\exp\{-\,\theta(h\,+\,T_{x-Y})\}] + o(h)\\&= e^{-\theta h}E[f(x-Y)] + o(h)\end{aligned}$$

其中 $o(h)$ 项来自到时间 h 过程击中 x 的可能性。将上式展开为关于 x 的泰勒级数得

$$\begin{aligned}f(x) &= e^{-\theta h}E[f(x) - Yf'(x) + \frac{Y^2}{2\,!}f''(x) + \cdots] + o(h)\\&= e^{-\theta h}[f(x) - \mu hf'(x) + \frac{h}{2}f''(x)] + o(h)\end{aligned}$$

现在利用 $e^{-\theta h} = 1 - \theta h + o(h)$ 给出

$$f(x) = f(x)(1-\theta h) - \mu h f'(x) + \frac{h}{2}f''(x) + o(h)$$

除以 h 并令 $h \to 0$ 得

$$\theta f(x) = -\mu f'(x) + \frac{1}{2}f''(x)$$

利用 $f(x) = e^{-cx}$

$$\theta e^{-cx} = \mu c e^{-cx} + \frac{c^2}{2}e^{-cx}$$

或

$$c^2 + 2\mu c - 2\theta = 0$$

由此可见

(6.4.10) $\quad c = -\mu + \sqrt{\mu^2 + 2\theta}$ 或 $c = -\mu - \sqrt{\mu^2 + 2\theta}$

由于 $c>0$，可见当 $\mu \geqslant 0$ 时

$$c = \sqrt{\mu^2 + 2\theta} - \mu$$

因此

(6.4.11)

$$E[e^{-\theta T_x}] = \exp\{-x(\sqrt{\mu^2 + 2\theta} - \mu)\}, \qquad \mu \geqslant 0$$

当 $\mu \geqslant 0$ 时，我们也能在(6.4.10)中定出 c 的恰当的值，首先注意到由于

$$E[\exp\{-\theta T_x\}] = \int_0^\infty e^{-\theta y} dF_{T_x}(y)$$

得

$$\lim_{\theta \to 0} e^{-cx} = P\{T_x < \infty\}$$

然而由等式(6.4.5)可见，当 $\mu<0$ 时

$$P\{T_x < \infty\} = e^{2\mu x}$$

这蕴含了这时 c 的恰当的值是

$$c = -\mu + \sqrt{\mu^2 + 2\theta}, \qquad \mu < 0^{*)}$$

综上所述我们有下列

命题 6.4.1

以 T_x 记漂移系数为 μ 的布朗运动过程击中 x 的时间．则对

$\theta>0, x>0$,

$$E[\exp\{-\theta T_x\}] = \exp\{-x(\sqrt{\mu^2+2\theta}-\mu)\}^{*)}$$

我们以研究最大值变量的极限平均值结束本节。具体地我们有下列

命题 6.4.2

若$\{X(t),\ t\geqslant 0\}$是漂移系数为μ，$\mu\geqslant 0$的布朗运动过程，则以概率 1

$$\lim_{t\to\infty}\frac{\max\limits_{0\leqslant s\leqslant t}X(s)}{t}=\mu$$

证明 令$T_0=0$，对$n<0$，以T_n记过程击中n的时刻。从平稳独立增量的假设得出，T_n-T_{n-1} $(n\geqslant 1)$是独立同分布的。因此我们可以把T_n看作一个更新过程中事件发生的时刻。以$N(t)$记到时刻t更新的次数，我们有

(6.4.12) $$N(t)\leqslant \max_{0\leqslant s\leqslant t}X(s)\leqslant N(t)+1$$

现在由例 6.4 (c) 我们有$ET_1=1/\mu$，从而由(6.4.12)及熟知的更新理论的结果$N(t)/t\to 1/ET_1$得命题结论。

6.5 向后与向前扩散方程

推导出微分方程是分析马尔可夫过程的一个有力技巧。有两个得到微分方程的一般技巧：向后与向前技巧。例如我们要求随机变量$X(t)$的密度。向后方法是对$X(h)$的值取条件，即回到时刻h的过程来看各种情况。向前方法是对$X(t-h)$取条件。

作为一个说明，考虑漂移系数为μ的布朗运动过程，以$p(x,t;y)$记$X(t)$在给定$X(0)=y$时的概率密度，即

$$p(x,t;y)=\lim_{\Delta x\to 0}P\{x<X(t)<x+\Delta x\,|\,X(0)=y\}/\Delta x$$

向后方法是对$X(h)$取条件。形式地把$p(x,t;y)$看作是概率，我们有

---

*）原书此处有误，已作修改。——译者注

$$p(x,t;y) = E[P\{X(t) = x | X(0) = y, X(h)\}]$$

现在

$$P\{X(t) = x | X(0) = y, X(h) = x_h\}$$
$$= P\{X(t-h) = x | X(0) = x_h\}$$

所以

$$p(x,t;y) = E[p(x,t-h;X(h))]$$

其中期望是对 $X(h)$ 取的，它是均值为 $\mu h + y$，方差为 h 的正态变量。假设上式右边能展开为$(x,t;y)$ 的泰勒级数，我们得到

$$\begin{aligned} p(x,t;y) = E\Big[&p(x,t;y) - h\frac{\partial}{\partial t}p(x,t;y) \\ &+ (X(h) - y)\frac{\partial}{\partial y}p(x,t;y) + \frac{h^2}{2}\cdot\frac{\partial^2}{\partial t^2}p(x,t;y) \\ &+ \frac{(X(h)-y)^2}{2}\frac{\partial^2}{\partial y^2}p(x,t;y) + \cdots\Big] \\ = p(x,t;y) &- h\frac{\partial}{\partial t}p(x,t;y) + \mu h\frac{\partial}{\partial y}p(x,t;y) \\ &+ \frac{h}{2}\frac{\partial^2}{\partial y^2}p(x,t;y) + o(h) \end{aligned}$$

除以 h 并让它趋于 0 得出

(6.5.1) $$\frac{1}{2}\frac{\partial^2}{\partial y^2}p(x,t;y) + \mu\frac{\partial}{\partial y}p(x,t;y) = \frac{\partial}{\partial t}p(x,t;y)$$

方程(6.5.1)称为向后扩散方程。

对 $X(t-h)$取条件得到向前方程。既然

$$\begin{aligned} P\{X(t) = x | X(0) = y, X(t-h) = a\} &\\ = P\{X(h) = x | X(0) = a\} &\\ = P\{W = x - a\} & \end{aligned}$$

其中 W 是均值为 μh，方差为 h 的正态随机度量。记它的密度为 f_W，则有

$$\begin{aligned} p(x,t;y) &= \int f_W(x-a)p(a,t-h;y)da \\ &= \int\Big[p(x,t;y) + (a-x)\frac{\partial}{\partial x}p(x,t;y) - h\frac{\partial}{\partial t}p(x,t;y) \end{aligned}$$

$$+\frac{(a-x)^2}{2}\frac{\partial^2}{\partial x^2}p(x,t;y)+\cdots]f_W(x-a)da$$

$$=p(x,t;y)-\mu h\frac{\partial}{\partial x}p(x,t;y)-h\frac{\partial}{\partial t}p(x,t;y)$$

$$+\frac{h}{2}\frac{\partial^2}{\partial x^2}p(x,t;y)+o(h)$$

除以 h 并令它趋于 0 得

(6.5.2) $$\frac{1}{2}\frac{\partial^2}{\partial x^2}p(x,t;y)=\mu\frac{\partial}{\partial x}p(x,t;y)+\frac{\partial}{\partial t}p(x,t;y)$$

方程（6.5.2）称为向前扩散方程。

6.6 利用柯尔莫哥洛夫方程求极限分布

向前微分方程方法最初用来求泊松过程 $N(t)$ 的分布，它对在众多的各种模型中求极限分布是十分有用的。这一方法利用时刻 t 系统状态的分布计算时刻 $t+h$ 状态的概率分布，并令 $t\to\infty$ 导出一个微分方程。现在我们说明它在几个模型中的用法，其中的第一个我们以前用其它方法研究过。

6.6.1 半马尔可夫过程

一个半马尔可夫过程是这样的过程：当它进入状态之后，在转移之前逗留一段随机时间，其分布为 H_i，均值为 μ_i。若在状态 i 停留的时间为 x，以概率 $P_{ij}(x),i,j\geqslant 0$，转移到状态 j。我们假设所有的分布 H_i 是连续的，定义其失效率函数 λ_i（t）为

$$\lambda_i(t)=h_i(t)/\overline{H}_i(t)$$

其中 h_i 为 H_i 的密度，因此在状态 i 停留了时间 t 的条件下，在接下来的 dt 时间内将转移的条件概率是 $\lambda_i(t)dt+o(dt)$。

若把任何时刻的“状态”看作（i，x），i 是现在的状态，x 是过程自进入 i 后已停留的时间，那么能把半马尔可夫过程分析为马尔可夫过程。令

$$p_t(i,x)=\lim_{h\to 0}\frac{P\left\{\begin{array}{c}\text{在时刻 } t \text{ 状态为 } i\text{，自进入 } i \text{ 以}\\ \text{来的时间在 } x-h \text{ 与 } x \text{ 之间}\end{array}\right\}}{h}$$

即 $p_t(i, x)$ 是时刻 t 状态为 (i, x) 的概率密度。

对 $x>0$，我们有

$$p_{t+h}(i,x+h)=p_t(i,x)(1-\lambda_i(x)h)+o(h) \tag{6.6.1}$$

因为要使时刻 $t+h$ 的状态为 $(i,x+h)$，过程必须在时刻 t 处于状态 (i,x)，且在接下来的 h 时间内没有发生转移。假设极限密度 $p(i,x)=\lim_{t\to\infty}p_t(i,x)$ 存在，在(6.6.1)中令 $t\to\infty$，我们有

$$\frac{p(i,x+h)-p(i,x)}{h}=-\lambda_i(x)p(i,x)+\frac{o(h)}{h}$$

令 $h\to 0$，有

$$\frac{d}{dx}p(i,x)=-\lambda_i(x)p(i,x)$$

除以 $p(i, x)$ 再积分得

$$\log\left(\frac{p(i,x)}{p(i,0)}\right)=-\int_0^x\lambda_i(y)dy$$

或

$$p(i,x)=p(i,0)\exp\left(-\int_0^x\lambda_i(y)dy\right)$$

等式（见第一章 1.6 节）

$$\overline{H}_i(x)=\exp\left(-\int_0^x\lambda_i(y)dy\right)$$

给出了

$$p(i,x)=p(i,0)\overline{H}_i(x) \tag{6.6.2}$$

此外，由于过程以概率强度 $\lambda_j(x)P_{ji}(x)$ 瞬时地从状态 (j,x) 转移到状态 $(i,0)$，我们也有

$$\begin{aligned}
p(i,0)&=\sum_j\int_x p(j,x)\lambda_j(x)P_{ji}(x)dx\\
&=\sum_j p(j,0)\int_x\overline{H}_j(x)\lambda_j(x)P_{ji}(x)dx \qquad (\text{由}(6.6.2))\\
&=\sum_j p(j,0)\int_x h_j(x)P_{ji}(x)dx
\end{aligned}$$

$\int_x h_j(x)P_{ji}(x)dx$ 正是过程进入状态 j 后接着进入状态 i 的概率。

把它记为P_{ji}，我们有

$$p(i,0)=\sum_j p(j,0)P_{ji}$$

现在若假设相继的状态所构成的转移概率为P_{ji}的马尔可夫链是遍历的，有极限概率$\pi_i(i\geqslant 0)$则由于$p(i,0)(i\geqslant 0)$满足平稳方程，故得对某个常数c及一切i有

(6.6.3) $$p(i,0)=c\pi_i$$

从（6.6.2）对x积分得

(6.6.4)
$$\begin{aligned}P\{\text{状态为}i\}&=\int p(i,x)dx\\&=p(i,0)\mu_i &&(\text{由}(6.6.2))\\&=c\pi_i\mu_i &&(\text{由}(6.6.3))\end{aligned}$$

由于$\sum_i P$｛状态为i｝$=1$，可见

$$c=\frac{1}{\sum_i \pi_i\mu_i}$$

所以，由（6.6.2）及（6.6.3）

(6.6.5) $$p(i,x)=\frac{\pi_i\mu_i}{\sum_i \pi_i\mu_i}\frac{\overline{H}_i(x)}{\mu_i}$$

注意到由（6.6.4）

(6.6.6) $$P\{\text{状态为}i\}=\frac{\pi_i\mu_i}{\sum_i\pi_i\mu_i}$$

及由（6.6.5）

$$P\{\text{处于此状态的时间}\leqslant y\mid\text{状态为}i\}=\int_0^y\frac{\overline{H}_i(y)}{\mu_i}ay$$

因此处于状态i的极限概率由（6.6.6）给出，与第四章的结果相吻合，在处于状态i的条件下，已在此状态中的时间服从H_i的平衡分布。

6.6.2 $M/G/1$ 排队系统

考虑$M/G/1$排队系统，来到是速率为λ的泊松过程，一个服务员，服务时间分布是G，假设G是连续的，失效率函数为$\lambda(t)$。这个模型可分析为马尔可夫过程，若让任一时刻的状态为(n,x)，

n 为系统中那时的人数，x 为正在被服务的顾客已服务的时间。

以 $p_t(n,x)$ 记在时刻 t 状态的密度，当 $n\geqslant 1$ 时有

$$(6.6.7)\qquad p_{t+x}(n,x+h) = p_t(n,x)(1-\lambda(x)h)(1-\lambda h) + p_t(n-1,x)\lambda h + o(h)$$

上式成立是因为时刻 $t+h$ 时的状态为 $(n,\ x+h)$ 若 (a) 时刻 t 时状态为 $(n,\ x)$，并在接下来的 h 时间内既无来到也无服务完毕，或 (b) 时刻 t 状态为 $(n-1,\ x)$ 并在接下来的 h 时间内有一个来到但无离去。

假定极限密度 $p(n,x)=\lim\limits_{t\to\infty}p_t(n,x)$ 存在，从(6.6.7)得

$$\frac{p(n,x+h)-p(n,x)}{h} = -(\lambda+\lambda(x))p(n,x)+\lambda p(n-1,x)+\frac{o(h)}{h}$$

令 $h\to 0$

$$(6.6.8)\qquad \frac{d}{dx}p(n,x) = -(\lambda+\lambda(x))p(n,x) + \lambda p(n-1,x), \qquad n\geqslant 1$$

现在定义母函数 $G(s,x)$ 为

$$G(s,x)=\sum_{n=1}^{\infty}s^n p(n,x)$$

求导数得

$$\begin{aligned}\frac{\partial}{\partial x}G(s,x) &= \sum_{n=1}^{\infty}s^n\frac{d}{dx}p(n,x)\\ &= \sum_{n=1}^{\infty}s^n[(-\lambda-\lambda(x))p(n,x) + \lambda p(n-1,x)] \qquad (\text{由}(6.6.8))\\ &= (\lambda s-\lambda-\lambda(x))G(s,x)\end{aligned}$$

两边除以 $G(s,x)$ 再积分得

$$\log\left(\frac{G(s,x)}{G(s,0)}\right) = (\lambda s-\lambda)x - \int_0^x\lambda(y)dy$$

或

$$G(s,x)=G(s,0)e^{-\lambda(1-s)x}\overline{G}(x) \tag{6.6.9}$$

上式中已利用了等式

$$\overline{G}(x)=\exp\left\{-\int_0^x\lambda(y)dy\right\}$$

为了得到 $G(s,0)$，注意到对于 $p(n,0)$，$n>0$ 的方程是

$$p(n,0)=\begin{cases}\int p(n+1,x)\lambda(x)dx, & n>1\\ \int p(n+1,x)\lambda(x)dx+P(0)\lambda, & n=1\end{cases}$$

其中

$$P(0)=P\{\text{系统是空的}\}$$

因此，

$$\sum_{n=1}^{\infty}s^{n+1}p(n,0)=\int\sum_{n=1}^{\infty}s^{n+1}p(n+1,x)\lambda(x)dx+s^2\lambda P(0)$$

或

$$\begin{aligned}sG(s,0)&=\int(G(s,x)-sp(1,x)\lambda(x)dx+s^2\lambda P(0)\\&=G(s,0)\int e^{-\lambda(1-s)x}g(x)dx\\&\quad-s\int p(1,x)\lambda(x)dx+s^2\lambda P(0)\end{aligned} \tag{6.6.10}$$

其中最后的等式是利用等式（6.6.9）的结果，为了计算（6.6.10）的右边的第二项，我们导出一个 $P(0)$ 的方程如下：

$$\begin{aligned}&P\{\text{在时刻 }t+h\text{ 系统空}\}\\&=P\{\text{在时刻 }t\text{ 系统空}\}(1-\lambda h)+\int\lambda(x)hp_t(1,x)dx+o(h)\end{aligned}$$

令 $t\to\infty$，然后 $h\to0$ 得

$$\lambda P(0)=\int\lambda(x)p(1,x)dx \tag{6.6.11}$$

将此代回(6.6.10)我们得到

$$sG(s,0)=G(s,0)\widetilde{G}(\lambda(1-s))-s\lambda(1-s)P(0)$$

或

(6.6.12) $$G(s,0)=\frac{s\lambda(1-s)P(0)}{\widetilde{G}(\lambda(1-s))-s}$$

其中 $\widetilde{G}(s)=\int e^{-xs}dG(x)$ 是服务时间分布的拉普拉斯变换。

为了得到系统中人数的边际概率母函数，积分(6.6.9)如下：

$$\begin{aligned}
&\sum_{n=1}^{\infty} s^n P\{\text{系统中有 } n \text{ 人}\}\\
&=\int_0^{\infty} G(s,x)dx\\
&=G(s,0)\int_0^{\infty} e^{-\lambda(1-s)x}\overline{G}(n)dx\\
&=G(s,0)\int_0^{\infty} e^{-\lambda(1-s)x}\int_x^{\infty} dG(y)dx\\
&=G(s,0)\int_0^{\infty}\int_0^{y} e^{-\lambda(1-s)x}dxdG(y)\\
&=\frac{G(s,0)}{\lambda(1-s)}\int_0^{\infty}(1-e^{-\lambda(1-s)y})dG(y)\\
&=\frac{G(s,0)(1-\widetilde{G}(\lambda(1-s)))}{\lambda(1-s)}
\end{aligned}$$

因此，由(6.6.12)

$$\begin{aligned}
&\sum_{n=1}^{\infty} s^n P\{\text{系统中有 } n \text{ 人}\}\\
&\qquad=P(0)+\frac{sP(0)(1-\widetilde{G}(\lambda(1-s)))}{\widetilde{G}(\lambda(1-s))-s}\\
&\qquad=\frac{P(0)(1-s)\widetilde{G}(\lambda(1-s))}{\widetilde{G}(\lambda(1-s))-s}
\end{aligned}$$

为了得到 $P(0)$ 的值，在上式中令 $s\to 1$，得

$$\begin{aligned}
1&=\sum_{n=1}^{\infty} P\{\text{系统中有 } n \text{ 人}\}\\
&=P(0)\lim_{s\to 1}\frac{(1-s)\widetilde{G}(\lambda(1-s))}{\widetilde{G}(\lambda(1-s))-s}
\end{aligned}$$

$$= P(0)\frac{\lim\limits_{s\to 1}\dfrac{d}{ds}(1-s)}{\lim\limits_{s\to 1}\dfrac{d}{ds}[\widetilde{G}(\lambda(1-s))-s]}$$

（由于 $\widetilde{G}(0)=1$，用洛必达法则）

$$= \frac{P(0)}{1-\lambda E[S]}$$

或

$$P(0) = 1-\lambda E[S]$$

其中 $E[S]=\int x dG(x)$ 是平均服务时间。

注记

(1) 我们也能试图从 $n=1$ 开始递推地得到函数 $p(n, x)$，并用 (6.6.8)作递推。例如，$n=1$ 时(6.6.8)中右边第二项消失了，因此结果为

$$\frac{d}{dx}p(1,x) = -(\lambda+\lambda(x))p(1,x)$$

解此方程得

$$p(1,x) = p(1,0)e^{-\lambda x}\overline{G}(x) \tag{6.6.13}$$

故

$$\int\lambda(x)p(1,x)dx = p(1,0)\int e^{-\lambda x}g(x)dx$$

利用(6.6.11)得

$$\lambda P(0) = p(1,0)\widetilde{G}(\lambda)$$

由于 $P(0)=1-\lambda E[S]$，可见

$$p(1,0) = \frac{\lambda(1-\lambda E[S])}{\widetilde{G}(\lambda)}$$

最后利用(6.6.13)我们有

$$p(1,x) = \frac{\lambda e^{-\lambda x}(1-\lambda E[S])\overline{G}(x)}{\widetilde{G}(\lambda)} \tag{6.6.14}$$

这公式可代入(6.6.8)，则 $p(2,x)$ 的微分方程至少在理论上可解。然后再能利用 $p(2,x)$ 去解 $p(3,x)$ 等等。

(2) 从(6.6.14)得出，在系统中只有一个顾客的条件下，顾客已被服务的时间的条件密度 $p(y|1)$ 为

$$p(y|1)=\frac{e^{-\lambda y}\overline{G}(y)}{\int e^{-\lambda y}\overline{G}(y)dy}$$

在 $G(y)=1-e^{-\mu y}$ 的特殊情形，我们有

$$p(y|1)=(\lambda+\mu)e^{-(\lambda+\mu)y}$$

因此，条件分布是参数为 $\lambda+\mu$ 的指数分布，不是平衡分布(它当然是参数为 μ 的指数分布)。

(3) 当然，在上述分析中为使极限分布存在需要条件 $\lambda E[S]<1$。

一般地，如果我们对计算一个马尔可夫过程 $\{X(t)\}$ 的极限概率分布有兴趣，那么利用向前方程是合适的方法。另一方面，如果我们对首达时间分布有兴趣，那么通常向后方程是最有价值的，也就是在这种问题中我们对最初 h 时间发生的事件取条件。

6.6.3 保险理论中的破产问题

假设到时刻 t 一家保险公司收到的赔偿要求的次数 $N(t)$ 是速率为 λ 的泊松过程。又假设相继的要求赔偿金额是独立的，有分布 G。若我们假设保险公司以每单位时间 1 元的常数速率收到保险金，则时刻 t 它的收支相抵可表示为

$$x+t-\sum_{i=1}^{N(t)}Y_i$$

其中 x 是公司的初始资本，Y_i $(i\geqslant 1)$ 是相继的赔偿要求。我们感兴趣的是公司永远有偿付能力的概率，它是初始资本 x 的函数，也就是我们要决定

$$R(x)=P\{x+t-\sum_{i=1}^{N(t)}Y_i>0,\text{对一切 }t\}$$

为了得到 $R(x)$ 的微分方程，我们用向后方法，对最初 h 时间发生的事件取条件。若无赔偿要求，则公司的资产为 $x+h$；而若有一个赔偿要求，则它们是 $x+h-Y$。因此

$$R(x) = R(x+h)(1-\lambda h) + E[R(x+h-Y)]\lambda h + o(h)$$

所以,

$$\frac{R(x+h)-R(x)}{h} = \lambda R(x+h) - \lambda E[R(x+h-Y)] + \frac{o(h)}{h}$$

令 $h \to 0$ 得

$$R'(x) = \lambda R(x) - \lambda E[R(x-Y)]$$

或

$$R'(x) = \lambda R(x) - \lambda \int_0^x R(x-y)dG(y)$$

有时能用此微分方程解出 R。

6.7 一个马尔可夫发射噪声过程

假设震动的发生遵循一个速率为 λ 的泊松过程。与第 i 次震动相联系有一个随机变量 $X_i(i \geqslant 1)$它表示震动的"值"。假定这些值是可叠加的,且它们以一个决定性的指数速率随时间而衰减。我们记

$N(t)$, 到时刻 t 震动的次数,

X_i, 第 i 次震动的值,

S_i, 第 i 次震动的时间,

则时刻 t 总的震动值, 记为 $X(t)$, 能表为

$$X(t) = \sum_{i=0}^{N(t)} X_i e^{-\alpha(t-S_i)}$$

其中 α 是常数, 它决定了指数衰减率。

当假设诸 X_i $(i \geqslant 1)$ 独立同分布, 且$\{X_i, i \geqslant 1\}$与泊松过程$\{N(t), t \geqslant 0\}$独立时,我们称$\{X(t), t \geqslant 0\}$为一个发射噪声过程。

发射噪声过程具有马尔可夫性, 给定现在的状态将来与过去条件独立。

我们能计算 $X(t)$的矩母函数,先对 $N(t)$取条件,然后用第二章的定理 2.3.1, 即给定 $N(t)=n$ 时不讲次序的来到时刻是独立

的(0,1)上的均匀分布随机变量。这给出

$$E[\exp\{sX(t)\}\,|\,N(t)=n]=E[\exp\{s\sum_{i=1}^{n}X_i e^{-\alpha(t-U_i)}\}]$$

其中$U_1,\cdots,U_n$是独立的$(0,t)$上均匀分布的随机变量。利用独立性将等式继续下去

$$\begin{aligned}E[\exp\{sX(t)\}\,|\,N(t)=n\,] &= (E[\exp\{sX_i e^{-\alpha(t-U_i)}\}])^n \\ &= \left[\int_0^t \phi(se^{-\alpha y})dy/t\right]^n \\ &\equiv \beta^n\end{aligned}$$

其中$\phi(u)=E[e^{uX}]$是X的矩母函数。因此

(6.7.1)
$$\begin{aligned}E[\exp\{sX(t)\}\,] &= \sum_{n=0}^{\infty}\beta^n e^{-\lambda t}\frac{(\lambda t)^n}{n!} \\ &= e^{-\lambda t}e^{\lambda t\beta} \\ &= \exp\{\lambda\int_0^t[\phi(se^{-\alpha y})-1]dy\}\end{aligned}$$

对上式求导可得$X(t)$的矩，我们留给读者去验证

(6.7.2)
$$\begin{aligned}E[X(t)] &= \lambda E[x](1-e^{-\alpha t})/\alpha \\ \mathrm{Var}[X(t)] &= \lambda E[X^2](1-e^{-2\alpha t})/2\alpha\end{aligned}$$

为了得到$\mathrm{Cov}(X(t),X(t+s))$，我们利用表示式

$$X(t+s)=e^{-\alpha s}X(t)+\overline{X}(s)$$

其中$\overline{X}(s)$与$X(s)$同分布，与$X(t)$独立，即$\overline{X}(s)$是在$(t,t+s)$中发生的事件在时刻$t+s$的作用。因此

$$\begin{aligned}\mathrm{Cov}(X(t),X(t+s)\,) &= e^{-\alpha s}\mathrm{Var}(X(t)) \\ &= e^{-\alpha s}\lambda E[X^2](1-e^{-2\alpha t})/2\alpha\end{aligned}$$

在(6.7.1)中令$t\to\infty$能得到$X(t)$的极限分布为

$$\lim_{t\to\infty}E[\exp\{sX(t)\}]=\exp\{\lambda\int_0^\infty[\phi(se^{-\alpha y})-1]dy\}$$

现在考虑X_i是参数为θ的指数随机变量的特殊情形。因此

$$\phi(u)=\frac{\theta}{\theta-u}$$

所以，此时

$$\lim_{t\to\infty} E[\exp\{sX(t)\}] = \exp\left\{\lambda\int_0^\infty \left(\frac{\theta}{\theta - se^{-\alpha y}} - 1\right) dy\right\}$$
$$= \exp\left\{\frac{\lambda}{\alpha}\int_0^s \frac{dx}{\theta - x}\right\}$$
$$= \left(\frac{\theta}{\theta - s}\right)^{\lambda/\alpha}$$

但 $(\theta/(\theta-s))^{\lambda/\alpha}$ 是参数为 λ/α 及 θ 的 Γ-随机变量的矩母函数。因此，当 X_i 是参数为 θ 的指数变量时，$X(t)$ 的极限密度为

$$(6.7.3)\qquad f(y) = \frac{\theta e^{-\theta y}(\theta y)^{\lambda/\alpha - 1}}{\Gamma(\lambda/\alpha)}, \qquad 0 < y < \infty$$

在本节的余下部分我们就假设 X_i 是参数为 θ 的指数变量，且过程处于稳态之下。对于后一要求，我们能设想 $X(0)$ 按照分布(6.7.3)选取，或者(还更好些)认为过程从 $t=-\infty$ 起始。

假设 $X(t)=y$。一个有趣的计算是决定自最近一次增加以来的时间的分布，即 t 以前最近一次泊松事件以来的时间的分布，称此随机变量为 $A(t)$，我们有

(6.7.4)

$$P\{A(t) > s \mid X(t) = y\}$$
$$= \lim_{h\to 0} P\{A(t) > s \mid y < X(t) < y + h\}$$
$$= \lim_{h\to 0} \frac{P\{ye^{\alpha s} < X(t-s) < (y+h)e^{\alpha s}, \text{在}(t-s,t)\text{ 中无事件}\}}{P\{y < X(t) < y + h\}}$$
$$= \lim_{h\to 0} \frac{f(ye^{\alpha s})e^{\alpha s}he^{-\lambda s} + o(h)}{f(y)h + o(h)}$$
$$= \exp\{-\theta y(e^{\alpha s} - 1)\}$$

应该指出，我们已经利用了过程处于稳态的假设而断定 $X(t)$ 与 $X(t-s)$ 的分布都由(6.7.3)给出。从(6.7.4)可见，给定 $X(t)=y$，$A(t)$ 的条件失效率函数，记为 $\lambda(s\mid y)$，是

$$\lambda(s\mid y) = \frac{\dfrac{d}{ds}P\{A(t) \leqslant s \mid X(t) = y\}}{P\{A(t) > s \mid X(t) = y\}}$$
$$= \theta\alpha ye^{\alpha s}$$

由此可见，朝着最近一次事件的逆向失效率在时刻 t 以 $\theta\alpha y$ 的速

率开始（即 $\lambda(0|y)=\theta\alpha y$），按时间逆向指数地增加直至下一个事件发生。应该指出，这显然不同于从时刻 t 至下一个事件的时间（它当然是与 $X(t)$ 独立的，参数为 λ 的随机变量）。

6.8 平稳过程

随机过程 $\{X(t),t\geqslant 0\}$ 称为平稳过程，若对一切 $n,s,t_1,\cdots,t_n$ 随机向量 $X(t_1),\cdots,X(t_n)$ 与 $X(t_1+s),\cdots,X(t_n+s)$ 有相同的联合分布。换句话说，过程是平稳的，若取任一固定点作为原点，随后的过程有相同的概率分布率。一些平稳过程的例子是：

(i)遍历的连续时间马尔可夫链 $\{X(t),t\geqslant 0\}$，当

$$P\{X(0)=j\}=P_j,\qquad j\geqslant 0$$

时，其中 $\{P_j,j\geqslant 0\}$ 是平稳分布。

(ii) $\{X(t),t\geqslant 0\}$，$X(t)$ 是平衡更新过程在时刻 t 的年龄。

(iii) $\{X(t),t\geqslant 0\}$，$X(t)=N(t+L)-N(t)$，$t\geqslant 0$，其中 $L>0$ 是固定常数，$\{N(t),t\geqslant 0\}$ 是速率为 λ 的泊松过程。

上述前两个过程是平稳的原因相同：它们是马尔可夫过程，初始状态按照极限状态分布选取（故它们可看作早已运行了无穷时间的遍历马尔可夫过程）。第三个例子（其中 $X(t)$ 表示泊松过程在 t 至 $t+L$ 之间发生的事件数）是平稳的得自泊松过程的平稳独立增量性假设。

过程为平稳的条件是相当严格的，所以我们定义过程 $\{X(t),\ t\geqslant 0\}$ 是二阶平稳的或协方差平稳的，若 $E[X(t)]=c$ 及 $\mathrm{Cov}(X(t),X(t+s))$ 与 t 无关。也就是过程是二阶平稳的（有时在文献中可看到的另一个名称是弱平稳的），若 $X(t)$ 的前二阶矩对一切 t 相同，$X(s)$ 与 $X(t)$ 之间的协方差只依赖于 $|t-s|$。对二阶平稳过程，令

$$R(s)=\mathrm{Cov}(X(t),X(s+t))$$

由于高斯过程的有限维分布（是多元正态分布）由它们的均值及协

方差决定，由此得出二阶平稳高斯过程是平稳的。然而有许多二阶平稳过程不平稳的例子。

例 6.8 (a)　自回归过程。设 Z_0，Z_1，Z_2，…是不相关的随机变量，$E[Z_n]=0$，$n\geqslant 0$，且

$$\mathrm{Var}(Z_n)=\begin{cases}\sigma^2/(1-\lambda^2), & n=0\\ \sigma^2, & n\geqslant 1\end{cases}$$

其中 $\lambda^2<1$。定义

(6.8.1)　　$X_0=Z_0$

(6.8.2)　　$X_n=\lambda X_{n-1}+Z_n,\qquad n\geqslant 1$

过程 $\{X_n,\ n\geqslant 0\}$ 称为一阶自回归过程。时刻 n 的状态 (X_n) 是时刻 $n-1$ 的状态的常数倍加上一个随机误差项(Z_n)。将(6.8.2)叠代得

$$\begin{aligned}X_n&=\lambda(\lambda X_{n-2}+Z_{n-1})+Z_n\\&=\lambda^2X_{n-2}+\lambda Z_{n-1}+Z_n\\&\ \ \vdots\\&=\sum_{i=0}^{n}\lambda^{n-i}Z_i\end{aligned}$$

所以

$$\begin{aligned}\mathrm{Cov}(X_n,X_{n+m})&=\mathrm{Cov}\Big(\sum_{i=0}^{n}\lambda^{n-i}Z_i,\sum_{i=0}^{n+m}\lambda^{n+m-i}Z_i\Big)\\&=\sum_{i=0}^{n}\lambda^{n-i}\lambda^{n+m-i}\mathrm{Cov}(Z_i,Z_i)\\&=\sigma^2\lambda^{2n+m}\Big(\frac{1}{1-\lambda^2}+\sum_{i=1}^{n}\lambda^{-2i}\Big)\\&=\frac{\sigma^2\lambda^m}{1-\lambda^2}\end{aligned}$$

上面计算中利用了 $i\neq j$ 时 Z_i 与 Z_j 不相关的事实。由于 $E[X_n]=0$，可见 $\{X_n,\ n\geqslant 0\}$ 是弱平稳的(对离散时间过程的定义明显地与连续时间过程的定义相似)。

例 6.8 (b)　动平均过程。设 W_0，W_1，W_2，…是不相关的，$E[W_n]=\mu$，$\mathrm{Var}(W_n)=\sigma^2$，$n\geqslant 0$，对某个正整数 k 定义

$$X_n=\frac{W_n+W_{n-1}+\cdots+W_{n-k}}{k+1},\qquad n\geqslant k$$

过程 $\{X_n,\ n\geqslant k\}$ 称为动平均过程，在每个时刻它是诸 W 的最近的 $k+1$ 个

值的算术平均。利用W_n，$n\geqslant 0$，互不相关可见

$$\mathrm{Cov}(X_n,X_{n+m})=\begin{cases}\dfrac{(k+1-m)\sigma^2}{(k+1)^2}, & 若\ 0\leqslant m\leqslant k\\ 0\qquad\qquad\quad, & 若\ m>k\end{cases}$$

因此，$\{X_n,\ n\geqslant k\}$是二阶平稳过程。

设$\{X_n,n\geqslant 1\}$是二阶平稳过程，$E[X_n]=\mu$。一个重要的问题是什么时候$\overline{X}_n\equiv\sum\limits_{i=1}^{n}X_i/n$收敛于$\mu$。下列命题证明，$E[(\overline{X}_n-\mu)^2]\to 0$当且仅当$\sum\limits_{i=1}^{n}R(i)/n\to 0$。即$\overline{X}_n$与$\mu$之差的平方的均值收敛于0当且仅当$R(i)$的极限平均值收敛于0。

命题 6.8.1

设$\{X_n,n\geqslant 1\}$是二阶平稳过程，有均值μ及协方差函数$R(i)=\mathrm{Cov}(X_n,X_{n+i})$，令$\overline{X}_n=\sum\limits_{i=1}^{n}X_i/n$。则

$$\lim_{n\to\infty}E[\overline{X}^n-\mu)^2]=0$$

当且仅当

$$\lim_{n\to\infty}\sum_{i=1}^{n}\frac{R(i)}{n}=0$$

证明 令$Y_i=X_i-\mu,\overline{Y}_n=\sum\limits_{i=1}^{n}Y_i/n$，假设$\sum\limits_{i=1}^{n}R(i)/n\to 0$。我们要证明这蕴含$E[\overline{Y}_n^2]\to 0$。现在有

$$\begin{aligned}E[\overline{Y}_n^2]&=\frac{1}{n^2}E\Big[\sum_{i=1}^{n}Y_i^2+2\sum_{i<j}\sum_{\leqslant n}Y_iY_j\Big]\\&=\frac{R(0)}{n}+\frac{2\sum\limits_{i<j}\sum\limits_{\leqslant n}R(j-i)}{n^2}\end{aligned}$$

我们留给读者去验证，当$\sum\limits_{i=1}^{n}R(i)/n\to 0$时，上式右边趋于0。

另一方面，假设$E[\overline{Y}_n^2]\to 0$，则

$$\Big(\sum_{i=0}^{n-1}\frac{R(i)}{n}\Big)^2=\Big[\frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}(Y_1,Y_i)\Big]^2$$

$$= [\mathrm{Cov}(Y_1, \overline{Y}_n)]^2$$
$$= [E(Y_1\overline{Y}_n)]^2$$
$$\leqslant E[Y_1^2]E[\overline{Y}_n^2]$$

这证明了，$n\to\infty$时$\sum_{i=0}^{n-1}R(i)/n\to 0$。读者应注意到上面利用了柯西—许瓦兹不等式：对随机变量 X 与 Y，$(E[XY])^2\leqslant E[X^2]E[Y^2]$（见习题 26）。

习 题

在习题 1，2 与 3 中，设$\{X(t),t\geqslant 0\}$为布朗运动过程。

1. 令 $Y(t) = tX(1/t)$
   (a)$Y(t)$的分布是什么？
   (b) 计算 $\mathrm{Cov}(Y(s),Y(t))$
   (c) 论证$\{Y(t),t\geqslant 0\}$也是布朗运动。
   (d) 令 $T = \inf\{t > 0; X(t) = 0\}$
   利用(c) 给出 $P\{T = 0\} = 1$ 的论证。
2. 令 $W(t) = X(a^2t)/a, a > 0$。验证$\{W(t),t\geqslant 0\}$也是布朗运动。
3. 计算给定 $X(t_1) = A, X(t_2) = B$ 时 $X(s)$ 的条件分布，其中 $t_1 < s < t_2$。
4. 设$\{Z(t),0\leqslant t\leqslant 1\}$为布朗桥过程。证明：若 $X(t) = (t+1)Z(t/(t+1))$，则$\{X(t),t\geqslant 0\}$是布朗运动过程。
5. 随机过程$\{X(t),t\geqslant 0\}$称为平稳的，若对一切 $n,a,t_1,\cdots,t_n$，$X(t_1),\cdots,X(t_n)$ 与 $X(t_1+a),\cdots,X(t_n+a)$ 有相同的联合分布。
   (a) 证明：高斯过程为平稳的充要条件是$\mathrm{Cov}(X(s),X(t))$只依赖于$t-s$，$s\leqslant t$，及 $E[X(t)] = c$。
   (b) 设$\{X(t),t\geqslant 0\}$为布朗运动，定义 $V(t) = e^{-\alpha t/2}X(e^{\alpha t})$。证明：$\{V(t),t\geqslant 0\}$是平稳高斯过程。它称为奥恩斯坦 — 乌伦佩克过程。
6. 设$\{X(t),t\geqslant 0\}$为一个生灭过程，它允许取负值，有常数生灭率 $\lambda_n\equiv\lambda,\mu_n\equiv\mu, n = 0, \pm 1, \pm 2,\cdots$。定义 μ 与 c 为 λ 的函数，使得 $\lambda\to\infty$ 时$\{cX(t),t\geqslant u\}$收敛于布朗运动。

在习题 7 到 12 中，设$\{X(t),t\geqslant 0\}$为布朗运动。

7. 求下列变量的分布：
   (a)$|X(t)|$

(b) $|\min_{0\leqslant s\leqslant t} X(s)|$

(c) $\max_{0\leqslant s\leqslant t} X(s) - X(t)$

8. 设 $X(1)=B$。按命题6.1.1的方式，刻划 $X(1)=B$ 的条件下的 $\{X(t), 0\leqslant t\leqslant 1\}$。

9. 令 $M(t)=\max_{0\leqslant s\leqslant t} X(s)$，证明：

$$P\{M(t)>a \mid M(t)=X(t)\}=e^{-a^2/2t}, \quad a>0$$

10. 计算布朗运动首次击中 n 的时间 T_x 的密度函数。

11. 以 T_1 记小于 t 的 $X(t)$ 的最大零点，以 T_2 记大于 t 的最小零点。证明：

(a) $P\{T_2 < s\} = (2/\pi)\arccos\sqrt{t/s}, \quad s>t$

(b) $P\{T_1 < s, T_2 > y\} = (2/\pi)\arcsin\sqrt{s/y}, \quad s<t<y$

12. 验证(6.3.4)中给出的 $|X(t)|$ 的均值与方差的公式。

13. 对漂移系数为 μ 的布朗运动证明：对 $x>0$

$$P\{\max_{0\leqslant s\leqslant h} |X(s)|>x\}=o(h)$$

14. 以 T_x 记布朗运动首次击中 x 的时间。计算：

$$P\{T_1<T_{-1}<T_2\}$$

15. 对漂移系数为 μ 的布朗运动，令

$$f(x)=E\,[\text{首次击中 } A \text{ 或} -B \text{ 的时间} \mid X_0=x]$$

其中 $A>0, B>0, -B<x<A$。

(a) 导出 $f(x)$ 的微分方程。

(b) 解此方程。

(c) 用随机游动取极限的论证方法（见第四章习题17）验证(b)中的解。

16. 以 T_a 记漂移系数为 μ 的布朗运动击中 a 的时间。

(a) 导出 $f(a,t)\equiv P\{T_a\leqslant t\}$ 满足的微分方程。

(b) 对 $\mu>0$，令 $g(x)=\mathrm{V_{ar}}(T_x)$，对 $g(x), x>0$，导出微分方程。

(c)(b)中的 $g(x), g(y)$ 与 $g(x+y)$ 之间有什么关系？

(d) 解出 $g(x)$。

(e) 通过对(6.4.11)求导验证你的解。

17. 在例6.4(b)中，假设 $X_0=x$ 及选手Ⅰ有加倍权，计算这种情况下Ⅰ的平均赢得。

18. 设 $\{X(t), t\geqslant 0\}$ 是不允许取负值的漂移系数为 $\mu(\mu<0)$ 的布朗运动。求 $X(t)$ 的极限分布。

19. 考虑有反射壁 $-B$ 及 A 的布朗运动，$A>0, B>0$。以 $p_t(x)$ 记 $X(t)$ 的密度函数。

(a) 导出 $p_t(x)$ 满足的微分方程。

(b) 求 $p(x)=\lim\limits_{t\to\infty}p_t(x)$。

20. 证明：对漂移系数为 μ 的布朗运动，概率为 1，$t\to\infty$ 时

$$\frac{X(t)}{t}\to\mu$$

21. 证明：

$$p(x,t;y)\equiv\frac{1}{\sqrt{2\pi t}}e^{-(x-y-\mu t)^2/2t}$$

满足向后与向前扩散方程(6.5.1)与(6.5.2)。

22. 验证等式(6.7.2)。

23. 验证当 $\{N(t)\}$ 是泊松过程时 $\{X(t)=N(t+L)-N(t), t\geqslant 0\}$ 是平稳的。

24. 设 U 服从 $(-\pi,\pi)$ 上均匀分布，令 $X_n=\cos(nU)$。利用三角恒等式

$$\cos x\cos y=\frac{1}{2}[\cos(x+y)+\cos(x-y)]$$

验证 $\{X_n, n\geqslant 1\}$ 是二阶平稳过程。

25. 证明：

$$\sum_{i=1}^{n}\frac{R(i)}{n}\to 0 \quad 蕴含 \quad \sum_{i<j}\sum_{\leqslant n}\frac{R(j-i)}{n^2}\to 0$$

从而完成命题 6.8.1 的证明。

26. 证明柯西－许瓦兹不等式：

$$(E[XY])^2\leqslant E[X^2]E[Y^2]$$

（提示：以不等式 $2|xy|\leqslant x^2+y^2$ 开始，然后以 $X/\sqrt{E[X^2]}$ 代 x，以 $Y/\sqrt{E[Y^2]}$ 代 y）

27. 对均值为 μ 且 $\sum\limits_{i=0}^{n-1}R(i)/n\to 0$ 的二阶平稳过程，证明：对任意 $\varepsilon>0, n\to\infty$ 时

$$\sum_{i=0}^{n-1}P\{|\overline{X}_n-\mu|>\varepsilon\}\to 0$$

参考文献

对经验分布函数收敛于布朗桥过程的严格证明有兴趣的学生(具备高等数学水平)应参阅文献1。文献5中给出了布朗桥对blackjack(又称“21点”的一种牌戏)的一个应用。文献3与6都对布朗运动有很好的处理。在它们的许多计算中，前者强调微分方程方法，而后者强调鞅方法，例6.4.2取自文献7，6.7节来自文献8。关于平稳过程的补充材料，有兴趣的读者应参阅文献2与4。

1. P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.
2. G. Box and G. Jenkins, *Time Series Analysis—Forecasting and Control*, Holden-Day, San Francisco, 1970.
3. D. R. Cox and H. D. Miller, *Theory of Stochastic Processes*, Methuen, London, 1965.
4. H. Cramer and M. Leadbetter, *Stationary and Related Stochastic Processes*, Wiley, New York, 1966.
5. G. Gottlieb, "An Analytic Derivation of Win Rate at Blackjack." *Management Science* (to appear).
6. S. Karlin and H. Taylor, *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York, 1975.
7. E. Keeler and J. Spencer, "Optimal Doubling in Backgammon," *Operations Research*, 23, No. 6 (1975), p. 1063.
8. G. Weiss, "A Shot Noise Process," Technical Report, Tel-Aviv University.

7 随机游动与鞅

引 言

设 X_1，X_2，…独立同分布（iid），且 $E[|X_i|]<\infty$。令 $S_0=0$，$S_n=\sum_{i=1}^{n}X_i$，$n\geqslant 1$。过程 $\{S_n,n\geqslant 0\}$ 称为随机游动过程。

为多种多样的现象建立模型时随机游动十分有用。例如，我们以前遇到过的简单随机游动 $(P\{X_i=1\}=p=1-P\{X_i=-1\})$，其中的 S_n 可解释为一个赌徒在 n 局赌博之后的赌金，在每一局中他赢或输一个单位的钱。我们也可用随机游动为更一般的赌博建立模型；例如，许多人相信，一家指定的上市公司的股票的相继的价格能建模为随机游动。我们将看到，在分析排队与破产系统中随机游动也是有用的。

在 7.1 节中我们给出一个对偶原理，它在求得有关随机游动的各种概率时十分有用。这节中最后一个例子处理 $G/G/1$ 排队系统，在分析它时我们被引导到去计算一个平均步子为负的随机游动迟早超过一个给定的常数的概率。

作为上述计算的前奏，在 7.2 节中我们介绍一类随机过程——鞅，它给出了公平赌博概念的数学定义。建立了鞅的一些性

质，并在 7.3 节中应用于随机游动。特别我们能近似地得到，而在某些情形能用明显表达式算出，一个随机游动迟早超过一个定值的概率。然后在 7.4 节中它们被用于 $G/G/1$ 排队系统与破产问题。

随机游动也能被看作更新过程的一种推广。因为若 X_i 被限制为非负随机变量，则 S_n 能被解释为一个更新过程的第 n 个事件的时刻。在 7.5 节中我们给出了布莱克威尔定理的一个推广，其中的 X_i 不要求是非负的，并指出了一个基于更新理论的结果的证明。

在本章的最后一节我们回到研究鞅这一类重要与有用的随机过程上来。在这节中我们给出了重要的鞅收敛定理，并用它证明强大数定律而结束本节。

7.1 随机游动中的对偶

设

$$S_n=\sum_{i=1}^{n}X_i,\qquad n\geqslant 1$$

为随机游动。在计算有关 $\{S_n, n\geqslant 1\}$ 的概率时，有一条对偶原理，尽管它是显然的，却十分有用。

对偶原理

$(X_1, X_2, \cdots, X_n)$与$(X_n,X_{n-1},\cdots,X_1)$有相同的联合分布。

由于 $X_i,i\geqslant 1$，是独立同分布的，因此立即可得此对偶原理成立。我们将在一系列的命题中阐明它的用处。

命题 7.11 说，若 $E[X]>0$，则随机游动在有限的平均步数之后变为正的。

命题 7.1.1

设 $X_1, X_2, \cdots$是独立同分布随机变量，且 $E[X]>0$。若

$$N=\min\{n: X_1+\cdots+X_n>0\}$$

则

$$E[N]<\infty$$

证明

$$\begin{aligned}E[N]&=\sum_{n=0}^{\infty}P\{N>n\}\\&=\sum_{n=0}^{\infty}P\{X_1\leqslant 0,X_1+X_2\leqslant 0,\cdots,X_1+\cdots+X_n\leqslant 0\}\\&=\sum_{n=0}^{\infty}P(X_n\leqslant 0,X_n+X_{n-1}\leqslant 0,\cdots,X_n+\cdots+X_1\leqslant 0)\end{aligned}$$

其中最后的等式得自对偶原理，所以

(7.1.1) $$E[N]=\sum_{n=0}^{\infty}P\{S_n\leqslant S_{n-1},S_n\leqslant S_{n-2},\cdots,S_n\leqslant 0\}$$

我们说一个更新在时刻 n 发生，如果 $S_n\leqslant S_{n-1}$，$S_n\leqslant S_{n-2}$，…，$S_n\leqslant 0$；即每当随机游动达到最低时一个更新发生。(略为想一想应使我们确信，相继的更新间的时间确实是独立同分布的。) 因此，由等式(7.1.1)，

$$\begin{aligned}E[N]&=\sum_{n=0}^{\infty}P\{\text{在时刻 } n \text{ 发生更新}\}\\&=1+E[\text{发生的更新数}]\end{aligned}$$

现在由于 $E[X]>0$，由强大数定律得 $S_n\rightarrow\infty$，所以发生的更新数(以概率1)是有限的。但是发生的更新数或者是无穷的，如果 $F(\infty)$(相继的更新间的时间为有限的概率)等于1；或者服从均值有限的几何分布，如果$F(\infty)<1$。由此

$$E\,[\text{发生的更新数}]<\infty$$

从而

$$E[N]<\infty$$

下一个命题讨论随机游动取新的值的平均速率问题。我们定义 R_n，称为$(S_0,S_1,\cdots,S_n)$的变程，如下。

定义 R_n 是$(S_0,S_1,\cdots S_n)$中的不相同的值的个数。

命题 7.1.2

$$\lim_{n\to\infty}\frac{E[R_n]}{n}=P\{\text{随机游动永不回到 }0\}$$

证明 令

$$I_k=\begin{cases}1, & 若 S_k\neq S_{k-1},\ S_k\neq S_{k-2},\ \cdots,\ S_k\neq S_0\\ 0, & 其它\end{cases}$$

则

$$R_n=1+\sum_{k=1}^{n}I_k$$

从而

$$\begin{aligned}E[R_n]&=1+\sum_{k=1}^{n}P\{I_k=1\}\\ &=1+\sum_{k=1}^{n}P\{S_k\neq S_{k-1},S_k\neq S_{k-2},\cdots,S_k\neq 0\}\\ &=1+\sum_{k=1}^{n}P\{X_k\neq 0,X_k+X_{k-1}\neq 0,\cdots,X_k+X_{k-1}+\cdots+X_1\neq 0\}\\ &=1+\sum_{k=1}^{n}P\{X_1\neq 0,X_1+X_2\neq 0,\cdots,X_1+\cdots+X_k\neq 0\}\end{aligned}$$

其中最后一个等式得自对偶原理。因此

$$\begin{aligned}E[R_n]&=1+\sum_{k=1}^{n}P\{S_1\neq 0,S_2\neq 0,\cdots,S_k\neq 0\}\\ &=\sum_{k=0}^{n}P\{T>k\}\end{aligned}\tag{7.1.2}$$

其中 T 是首次回到 0 的时刻，现在 $k\to\infty$ 时

$$P\{T>k\}\to P\{T=\infty\}=P\{不回到\ 0\}$$

故由(7.1.2)可见

$$E[R_n]/n\to P(不回到\ 0)$$

例 7.1(a) 简单随机游动。 在简单随机游动中 $P\{X_i=1\}=p=1-P\{X_i=-1\}$。现在当 $p=\frac{1}{2}$(对称简单随机游动)时，随机游动是常返的，因此，$p=\frac{1}{2}$时

$$P\{不回到\ 0\}=0$$

从而 $p>\frac{1}{2}$时

$$E[R_n/n]\to 0$$

$p=\frac{1}{2}$时，令 $\alpha=P\{回到\ 0|X_1=1\}$。由于 $P\{回到\ 0|X_1=-1\}=1$(为什么?)我们有

$$P\{回到\ 0\}=\alpha p+1-p$$

也对 X_2 取条件得

$$\alpha=\alpha^2 p+1-p$$

或等价地

$$(\alpha-1)(\alpha p-1-p)=0$$

由于非常返性，$\alpha<1$，可见

$$\alpha=(1-p)/p$$

所以，$p>\frac{1}{2}$ 时

$$E[R_n/n]\to 2p-1$$

类似地，$p\leqslant\frac{1}{2}$ 时

$$E[R_n/n]\to 2(1-p)-1$$

对偶的下一个应用是处理对称随机游动的。

命题 7.1.3

在对称简单随机游动中，对一切 $k\neq 0$，在回到原点之前到达状态 k 的平均次数为 1。

证明 对 $k>0$，以 Y 记首次回到原点之前到达状态 k 的次数，那么 Y 可表为

$$Y=\sum_{n=1}^{\infty} I_n$$

其中

$$I_n=\begin{cases}1, & \text{若在时刻 } n \text{ 到达状态 } k, n \text{ 之前未回到原点}\\ 0, & \text{其它}\end{cases}$$

或等价地

$$I_n=\begin{cases}1, & \text{若 } S_n>0, S_{n-1}>0, S_{n-2}>0, \cdots, S_1>0, S_n=k\\ 0, & \text{其它}\end{cases}$$

从而

$$\begin{aligned}E[Y]&=\sum_{n=1}^{\infty} P\{S_n>0,\ S_{n-1}>0,\ \cdots,\ S_1>0,\ S_n=k\}\\ &=\sum_{n=1}^{\infty} P\ \{X_n+\cdots+X_1>0,\ X_{n-1}+\cdots+X_1>0,\ \cdots,\\ &\qquad\qquad X_1>0,\ X_n+\cdots+X_1=k\}\\ &=\sum_{n=1}^{\infty} P\ \{X_1+\cdots+X_n>0,\ X_2+\cdots+X_n>0,\ \cdots,\end{aligned}$$

$$X_n>0,\ X_1+\cdots+X_n=k\}$$

其中最后一个等式得自对偶原理。因此

$$\begin{aligned}E[Y]&=\sum_{n=1}^{\infty}P\{S_n>0,\ S_n>S_1,\ \cdots,\ S_n>S_{n-1},\ S_n=k\}\\&=\sum_{n=1}^{\infty}P\{\text{对称随机游动在时刻 } n \text{ 首次击中 } k\}\\&=P\{\text{对称随机游动迟早击中 } k\}\\&=1 \qquad\qquad (\text{由常返性})\end{aligned}$$

证毕*。

最后将对偶原理应用于 $G/G/1$ 的排队模型。这是一个单服务员模型，顾客遵循一个更新过程来到，它有任意的来到间隔分布 F，服务时间分布为 G。记来到间隔时间为 X_1，X_2，…，服务时间为 Y_1，Y_2，…，第 n 个顾客在排队中的等待时间为 D_n。由于第 n 个顾客在系统中化费掉时间 D_n+Y_n，第 $n+1$ 个顾客在第 n 个顾客之后 X_{n+1} 时间来到，故(见图 7.1.1)

$$D_{n+1}=\begin{cases}D_n+Y_n-X_{n+1}, & \text{若 } D_n+Y_n\geqslant X_{n+1}\\ 0\qquad\qquad\quad, & \text{若 } D_n+Y_n<X_{n+1}\end{cases}$$

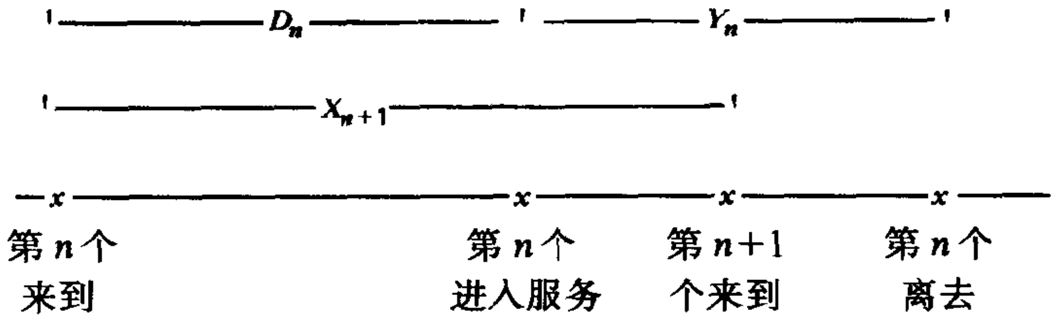
或等价地，令 $U_n=Y_n-X_{n+1}$，$n\geqslant 1$，

(7.1.3) $\qquad D_{n+1}=\max\ \{0,\ D_n+U_n\},\qquad n\geqslant 0$

叠代 (7.1.3) 得

$$\begin{aligned}D_{n+1}&=\max\ \{0,\ D_n+U_n\}\\&=\max\ \{0,\ U_n+\max\ \{0,\ D_{n-1}+U_{n-1}\}\}\\&=\max\ \{0,\ U_n,\ U_n+U_{n-1}+D_{n-1}\}\\&=\max\ \{0,\ U_n,\ U_n+U_{n-1}+\max\ \{0,\ U_{n-2}+D_{n-2}\}\}\\&=\max\ \{0,\ U_n,\ U_n+U_{n-1},\ U_n+U_{n-1}+U_{n-2}+D_{n-2}\}\\&\ \ \vdots\\&=\max\ \{0,\ U_n,\ U_n+U_{n-1},\ \cdots,\ U_n+U_{n-1}+\cdots+U_1\}\end{aligned}$$

* 读者应将此证明与第四章习题 31 中所勾画出的证明作比较。



或者

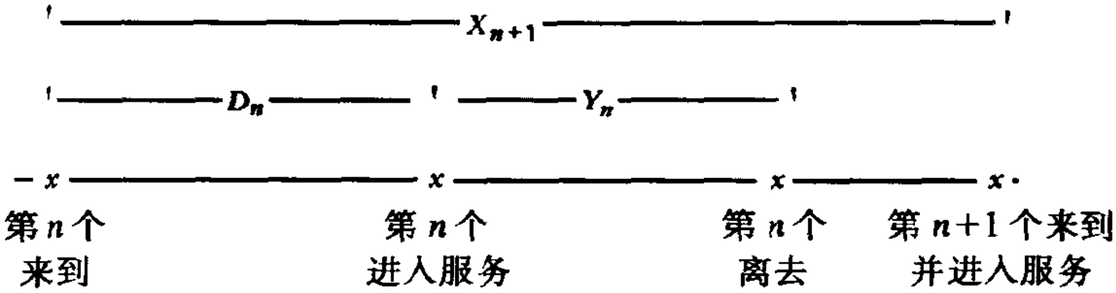


图 7.1.1　$D_{n+1}=\max\{D_n+Y_n-X_{n+1},0\}$

其中最后一步利用了 $D_1=0$ 的事实。因此，对 $c>0$，

$$P\{D_{n+1}\geqslant c\}=P\{\max(0,U_n,U_n+U_{n-1},\cdots,U_n+\cdots+U_1)\geqslant c\}$$
$$=P\{\max(0,U_1,U_1+U_2,\cdots,U_1+\cdots+U_n)\geqslant c\}$$

其中最后的等式得自对偶原理。从而我们有下列

命题 7.1.4

设 D_n 是 $G/G/1$ 排队系统中第 n 个顾客在排队中的等待时间，来到间隔时间为 $X_i(i\geqslant 1)$，服务时间为 $Y_i(i\geqslant 1)$，则对 $c>0$

(7.1.4)　　$P\{D_{n+1}\geqslant c\}=P\{$随机游动 $S_j,j\geqslant 1$,到时刻 n 越过 $c\}$

其中

$$S_j=\sum_{i=1}^{j}(Y_i-X_{i+1})$$

由命题 7.1.4，我们也看出 $P\{D_{n+1}\geqslant c\}$关于 n 非降。令

$$P\{D_\infty\geqslant c\}=\lim_{n\to\infty}P\{D_n\geqslant c\}$$

由(7.1.4)我们有

(7.1.5)　　$P\{D_\infty\geqslant c\}=P\{$随机游动 $S_j,j\geqslant 1$,迟早越过 $c\}$

若 $E[U]=E[Y]-E[X]$ 是正的，则由强大数定律，随机游动将趋于无穷，所以若 $E[Y]>E[X]$，对一切 c

$$P\{D_{\infty}\geqslant c\}=1$$

上述结论在 $E[Y]=E[X]$ 时也成立，从而仅在 $E[Y]<E[X]$ 时存在极限等待分布。这时为计算 $P\{D_{\infty}>c\}$ 我们需要计算一个平均变化为负的随机游动迟早越过一个常数的概率。然而，在考虑此问题之前，先引进鞅的概念是十分有用的。

7.1.1 关于可交换随机变量的一些注记

为了得到对偶原理并非必须假设随机变量 $X_1, \cdots, X_n$ 为独立同分布的。一个较弱的一般条件是这些随机变量是可交换的，我们说 $X_1, \cdots, X_n$ 是可交换的，如果对 $(1,2,\cdots,n)$ 的一切置换 $(i_1, \cdots, i_n)$，$X_{i_1}, \cdots, X_{i_n}$ 有相同的联合分布。

例 7.1(b) 从一个最初装有 n 个球(其中 k 个白球)的罐中不放回地随机取球。若令

$$X_i=\begin{cases}1, & \text{若第 } i \text{ 次取到白球}\\ 0, & \text{其它}\end{cases}$$

则 $X_1, \cdots, X_n$ 是可交换的，但不独立。

作为说明利用可交换性的一个例子，假设 X_1 与 X_2 是可交换的，$f(x)$ 与 $g(x)$ 为增函数，则对一切 x_1, x_2

$$(f(x_1)-f(x_2))(g(x_1)-g(x_2))\geqslant 0$$

这蕴含了

$$E[(f(X_1)-f(X_2))(g(X_1)-g(X_2)]\geqslant 0$$

但可交换性又蕴含了

$$E[f(X_1)g(X_1)]=E[f(X_2)g(X_2)]$$

$$E[f(X_1)g(X_2)]=E[f(X_2)g(X_1)]$$

将上列不等式展开即可见

$$E[f(X_1)g(X_1)]\geqslant E[f(X_1)g(X_2)]$$

特别若 X_1 与 X_2 独立，我们有下列

命题 7.1.5

若 f 与 g 都是递增函数，则

$$E[f(X)g(X)]\geqslant E[f(X)]E[g(X)]$$

随机变量的无穷序列 X_1，X_2，…称为可交换的，若每个有限子序列 X_1，…，X_n 是可交换的。

例 7.1（c） 设 Λ 为随机变量，分布为 G。假设在事件 $\Lambda=\lambda$ 的条件下，X_1，X_2，…为独立同分布，分布为 F_λ，即

$$P\{X_1\leqslant x_1,\ \cdots,\ X_n\leqslant x_n\mid\Lambda=\lambda\}=\prod_{i=1}^{n}F_\lambda(x_i)$$

随机变量 X_1，X_2，…是可交换的，因为

$$P\{X_1\leqslant x_1,\cdots,X_n\leqslant x_n\}=\int\prod_{i=1}^{n}F_\lambda(x_i)dG(\lambda)$$

关于$(x_1,\cdots,x_n)$是对称的，然而它们不是独立的。

有一个以德·费奈蒂定理著称的结果，它说每个可交换随机变量的无穷序列有例 7.1(c) 所规定的形式。我们将给出 X_i 是 0—1（即贝努里）随机变量时的证明。

*定理 7.1.6（德·费奈蒂定理） 对每一个可交换的取值 0 或 1 的随机变量 X_1，X_2，…的无穷序列，对应有一个$[0,1]$上的概率分布 G，使得对一切 $0\leqslant k\leqslant n$

$$P\{X_1=X_2=\cdots=X_k=1,\ X_{k+1}=\cdots=X_n=0\}=\int_0^1\lambda^k(1-\lambda)^{n-k}dG(\lambda)\tag{7.1.6}$$

证明 设 $m\geqslant n$。我们首先对

$$S_m=\sum_{i=1}^{m}X_i$$

取条件来计算上述概率。这导致

$$\begin{aligned}P\{X_1=\cdots=X_k=1,X_{k+1}=\cdots=X_n=0\}&=\sum_{j=0}^{m}P\{X_1=\cdots=X_k=1,X_{k+1}=\cdots=X_n=0\mid S_m=j\}P\{S_m=j\}\\&=\sum_{j=0}^{m}\frac{j(j-1)\cdots(j-k+1)(m-j)(m-j-1)\cdots(m-j-(n-k)+1)}{m(m-1)\cdots(m-n+1)}P\{S_m=j\}\end{aligned}\tag{7.1.7}$$

推得最后一个等式是因为在 $S_m=j$ 的条件下由可交换性可知，$X_1,\cdots,X_m$ 的

每一个由 j 个变量组成的子集都以同样的可能性是全由 1 组成的集合。

令 $Y_m=\frac{S_m}{m}$，则(7.1.7)可写成

(7.1.8) $$P\{X_1=\cdots=X_k=1,\ X_{k+1}=\cdots=X_n=0\}$$

$$=E\left[\frac{(mY_m)(mY_m-1)\cdots(mY_m-k+1)[m(1-Y_m)][m(1-Y_m)-1]\cdots[m(1-Y_m)-n+k+1]}{m(m-1)\cdots(m-n+1)}\right]$$

对充分大的 m，上式大致等于 $E[Y_m^k(1-Y_m)]^{n-k}$，只要令 $m\to\infty$ 就证得定理。事实上利用一个称为赫利定理的结果能够证明，对某个趋于∞的子列 m'，$Y_{m'}$ 的分布收敛于一个分布 G，而(7.1.8)将收敛于

$$E[Y_\infty^k(1-Y_\infty)^{n-k}]=\int_0^1 y^k(1-y)^{n-k}dG(y)$$

注记　对可交换随机变量的有限序列，德·费奈蒂定理不成立。例如，若在例 7.1(b)中取 $n=2,k=1$，则 $P\{X_1=1,X_2=0\}=P\{X_1=0,X_2=1\}=½$，它们不可能写成(7.1.6)的形式。

7.2 鞅

随机过程 $\{Z_n,n\geqslant1\}$ 称为鞅，如果对一切 n

$$E\{|Z_n|\}<\infty$$

及

(7.2.1) $$E[Z_{n+1}|Z_1,Z_2,\cdots,Z_n]=Z_n$$

鞅是公平赌博的一种推广。假若我们把 Z_n 解释为一个赌徒 n 局之后的赌金，那么(7.2.1)是说他在 $n+1$ 局之后的平均赌金等于他在 n 局之后的赌金，不论以前发生怎样的情况。

对(7.2.1)式取期望得

$$E[Z_{n+1}]=E[Z_n]$$

所以，对一切 n

$$E[Z_n]=E[Z_1]$$

鞅的几个例子　(1) 设 X_1，X_2，…为独立随机变量，均值为 0；令 $Z_n=\sum_{i=1}^{n}X_i$，则 $\{Z_n,n\geqslant1\}$ 是鞅，因为

$$E[Z_{n+1}|Z_1,\cdots,Z_n]=E[Z_n+X_{n+1}|Z_1,\cdots Z_n]$$

$$
\begin{aligned}
&=E[Z_n|Z_1,\cdots,Z_n]+E[X_{n+1}|Z_1,\cdots,Z_n]\\
&=Z_n+E[X_{n+1}]\\
&=Z_n
\end{aligned}
$$

(2) 若 X_1，X_2，…为独立随机变量，$E[X_i]=1$，则 $Z_n=\prod_{i=1}^{n}X_i$ 时，$\{Z_n, n\geqslant 1\}$是鞅。这是因为

$$
\begin{aligned}
E[Z_{n+1}|Z_1,\cdots,Z_n]&=E[Z_nX_{n+1}|Z_1,\cdots Z_n]\\
&=Z_nE[Z_{n+1}|Z_1,\cdots,Z_n]\\
&=Z_nE[X_{n+1}]\\
&=Z_n
\end{aligned}
$$

(3) 考虑一个分支过程（见第四章 4.5 节），以 X_n 记第 n 代的总量。若 m 是每个个体的平均后代数，则

$$Z_n=X_n/m^n$$

时，$\{Z_n, n\geqslant 1\}$是鞅。我们把证明留作练习。

定义 正整数值（可能取无穷值）的随机变量 N 称为过程 $\{Z_n,n\geqslant 1\}$的随机时间，如果事件$\{N=n\}$由随机变量 $Z_1,\cdots,Z_n$ 决定，也就是，知道 Z_1，…，Z_n 就告诉了我们是否有 $N=n$。若 $P\{N<\infty\}=1$，随机时间 N 称为停时。

设 N 是过程$\{Z_n,n\geqslant 1\}$的随机时间，令

$$\overline{Z}_n=\begin{cases}Z_n, & 若\ n\leqslant N\\ Z_N, & 若\ n>N\end{cases}$$

$\{\overline{Z}_n, n\geqslant 1\}$称为停止过程。

命题 7.2.1

若 N 是鞅$\{Z_n\}$的随机时间，则停止过程$\{\overline{Z}_n\}$也是鞅。

证明 令

$$I_n=\begin{cases}1, & 若\ N\geqslant n\\ 0, & 若\ N<n\end{cases}$$

也就是，若观察 Z_1，…，Z_{n-1}之后我们没有停止，则 $I_n=1$。我们说成立等式

(7.2.2) $$\overline{Z}_n=\overline{Z}_{n-1}+I_n(Z_n-Z_{n-1})$$

为了验证 (7.2.2) 考虑两种情况：

(i) $N \geqslant n$：这时 $\bar{Z}_n = Z_n$，$\bar{Z}_{n-1} = Z_{n-1}$，$I_n = 1$，故得(7.2.2)。

(ii) $N < n$：这时 $\bar{Z}_{n-1} = \bar{Z}_n = Z_N$，$I_n = 0$，故得(7.2.2)。

现在有

(7.2.3) $$E[\bar{Z}_n | Z_1, \cdots, Z_{n-1}] = E[\bar{Z}_{n-1} + I_n(Z_n - Z_{n-1}) | Z_1, \cdots Z_{n-1}]$$
$$= \bar{Z}_{n-1} + I_n E[Z_n - Z_{n-1} | Z_1, \cdots, Z_{n-1}]$$
$$= \bar{Z}_{n-1}$$

其中最后第二个等式成立是因为 $\bar{Z}_{n-1}$ 与 I_n 都由 $Z_1, \cdots, Z_{n-1}$ 决定，而最后一个等式成立是由于 $\{Z_n\}$ 为鞅。

我们必须证明 $E[\bar{Z}_n | \bar{Z}_1, \cdots, \bar{Z}_{n-1}] = \bar{Z}_{n-1}$。然而(7.2.3)已蕴含这个结论，因为若我们知道 Z_1，…，Z_{n-1} 的值。那么我们也知道 $\bar{Z}_1$，…，$\bar{Z}_{n-1}$ 的值。更正式地，我们有

$$E[\bar{Z}_n | \bar{Z}_1, \cdots, \bar{Z}_{n-1}]$$
$$= E[E[\bar{Z}_n | Z_1, \cdots, Z_{n-1}, \bar{Z}_1, \cdots, \bar{Z}_{n-1}] | \bar{Z}_1, \cdots, \bar{Z}_{n-1}]$$
$$= E[E[\bar{Z}_n | Z_1, \cdots, Z_{n-1}] | \bar{Z}_{11}, \cdots, \bar{Z}_{n-1}]$$
$$= E[\bar{Z}_{n-1} | \bar{Z}_1, \cdots, \bar{Z}_{n-1}] \qquad (\text{由}(7.3.3))$$
$$= \bar{Z}_{n-1}$$

由于停止过程也是鞅，$\bar{Z}_1 = Z_1$，故对一切 n

(7.2.4) $$E[\bar{Z}_n] = E[Z_1]$$

现在假设随机时间 N 是停时，即 $P\{N < \infty\} = 1$。由于

$$\bar{Z}_n = \begin{cases} Z_n, & \text{若 } n \leqslant N \\ Z_N, & \text{若 } m > N \end{cases}$$

故得 n 充分大时 $\bar{Z}_n$ 等于 Z_N，因此 $n \to \infty$ 时以概率 1

$$\bar{Z}_n \to Z_N$$

那么在 $n \to \infty$ 时

(7.2.5) $$E[\bar{Z}_n] \to E[Z_N]$$

是否成立呢？由于对一切 n，$E[\bar{Z}_n] = E[Z_1]$，(7.2.5)就是说

$$E[Z_N] = E[Z_1]$$

实际情况是在某些正则性条件下(7.2.5)确实成立。我们不加证明地叙述下列定理。

定理 7.2.2 若或者 (i) $\bar{Z}_n$ 一致有界，或者 (ii) N 有界，或

者（iii）$E[N]<\infty$，且存在一个 $M<\infty$ 使得

$$E[|Z_{n+1}-Z_n||Z_1,\cdots,Z_n]<M$$

则(7.2.5)成立，从而

$$E[Z_N]=E[Z_1]$$

定理 7.2.2 说，在一个公平的赌博中，如果赌徒使用一个停时去决定什么时候退出，那么他的最后的平均赌金等于他的初始的平均赌金。因此在均值的意义下，若定理 7.2.2 中的充分条件之一满足，那么成功的赌博系统是不可能有的。

系 7.2.3（瓦尔德等式）

若 X_i（$i\geqslant 1$）为 iid，$E[|X|]<\infty$，N 是 $X_1,X_2,\cdots$ 的停时，$E[N]<\infty$，则

$$E[\sum_{i=1}^{N}X_i]=E[N]E[X]$$

证明 令 $\mu=E[X]$。由于

$$Z_n=\sum_{i=1}^{n}(X_i-\mu)$$

是鞅，倘若定理 7.2.2 可应用，则得

$$E[Z_N]=E[Z_1]=0$$

但是

$$\begin{aligned}E[Z_N]&=E\Big[\sum_{i=1}^{N}(X_i-\mu)\Big]\\&=E\Big[\sum_{i=1}^{N}X_i-N\mu\Big]\\&=E\Big[\sum_{i=1}^{N}X_i\Big]-E[N]\mu\end{aligned}$$

为了证明定理 7.2.2 可应用，我们验证条件(iii)。现在 $Z_{n+1}-Z_n=X_{n+1}-\mu$，故

$$\begin{aligned}E[|Z_{n+1}-Z_n||Z_1,\cdots,Z_n]&=E[|(X_{n+1}-\mu|)|Z_1,\cdots,Z_n]\\&=E[|X_{n+1}-\mu|]\\&\leqslant E[|X|]+\mu\end{aligned}$$

例 7.2(a)　计算直到一种给定的花样出现的平均时间。假设序贯地观察一列独立同分布的离散随机变量，一天一个。直到某个给定的序列出现必

须观察的平均次数是多少呢？更具体地，设每次观察的结果分别以概率$\frac{1}{2}$，$\frac{1}{3}$及$\frac{1}{6}$为0，1及2，我们希望求直到游程020出现为止的平均时间。例如，若观察结果的序列是2，1，2，0，2，1，0，1，0，0，2，0，则所需的数N等于12。

为了计算$E[N]$，设想有一列赌徒在一个公平的赌场参赌，每人起初都有1元。第i个赌徒在第i天的开始赌起，以他的1元打赌那天的观察值将是0。如果他赢了(从而他有2元)，就以2元打赌下一个结果是2。如果这次赢了(从而他有12元)，就以全部12元打赌下一个结果是0。因此每个赌徒如在三次打赌中输掉任何一次，将输1元，但三次都赢了将赢23元。每天一开始又一个赌徒开始赌。以X_n记第n天结束时赌场的总赢得。由于所有的打赌都是公平的，故得$\{X_n, n\geqslant 1\}$是鞅，均值为0。以N记直到序列020出现的时间。现在第N天结束时赌徒1，…，$N-3$都输1元，赌徒$N-2$赢了23元，赌徒$N-1$输1元，赌徒N赢1元(因为第N天的观察结果是0)。所以

$$X_N = N-3-23+1-1 = N-26$$

又由于$E[X_N]=0$(容易验证定理7.2.2的条件(iii))可见

$$E[N]=26$$

用如上同样的方式，我们能计算直到任一给定的花样出现为止的平均时间。例如掷硬币直到HHTTHH (H—正面，T—反面) 出现的平均时间是$p^{-4}q^{-2}+p^{-2}+p^{-1}$，其中$p=P\{H\}=1-q$。

假设现在我们要计算一个给定的花样(譬如是花样A)比第二个花样(譬如是花样B)早出现的概率。为了计算这概率，先来考虑在一个给定的花样出现之后直到另一个花样出现的附加时间是有益的。例如在上面的例子中设$A=0, 2, 0$，$B=1, 0, 0, 2$，考虑$N_{A|B}$，B之后A出现所需的附加试验次数。也就是

$$N_{A|B}=\min\{k: 0,2,0 \text{ 是 } 1,0,0,2,X_1,X_2,\cdots,X_k \text{ 的连通子集}\}$$

为计算$E[N_{A|B}]$，再设想每一天一个赌徒开始打赌子序列$A=0, 2, 0$，在接下去的三个观察结果中出现。现在给定最初四个结果是1，0，0，2，故得赌徒1与2都输掉1元，赌徒3赢了11元(因为他赢了前二天的赌局)，赌徒4输掉1元。现在如前面一样，在第$4+N_{A|B}$天，赌场将赢$4+N_{A|B}-26=N_{A|B}-22$元。由于赌场的赢得构成鞅，在前四天的赌博中输掉8元，故得

$$E[N_{A|B}-22]=-8$$

以 N_A 记为了得到 A 所需的试验数，我们证明了

$$E[N_A]=26, \qquad E[N_{A|B}]=14$$

(作为一种核查，注意到由于 $A=0, 2, 0, B=1, 0, 0, 2$，必须有 $E[N_{A|B}]=1+\frac{1}{2}E[N_A]$。)类似地，我们能证明

$$E[N_B]=72, \qquad E[N_{B|A}]=72$$

为了计算 A 在 B 之前出现的概率 P_A，令 $M=\min(N_A, N_B)$，则

$$\begin{aligned}E[N_A]&=E[M]+E[N_A-M]\\&=E[M]+E[N_A-M|B\text{ 在 }A\text{ 之前}](1-P_A)\\&=E[M]+(1-P_A)E[N_{A|B}]\end{aligned}$$

类似地

$$E[N_B]=E[M]+P_AE[N_{B|A}]$$

解这些方程得

$$P_A=\frac{E[N_B]+E[N_{A|B}]-E[N_A]}{E[N_{B|A}]+E[N_{A|B}]}$$

$$E[M]=E[N_B]-E[N_{B|A}]\frac{E[N_B]+E[N_{A|B}]-E[N_A]}{E[N_{B|A}]+E[N_{A|B}]}$$

在我们所考虑的特殊情形

$$P_A=\frac{72+14-26}{14+72}=\frac{30}{43}$$

$$E[M]=72-\frac{30}{43}(72)=\frac{936}{43}$$

例 7.2(b) 假设我们要计算简单随机游动直到击中 i，$i>0$，为止的平均时间，其中 $p>\frac{1}{2}$。为此以 N 记这段时间。由于

$$\sum_{j=1}^{N}X_j=i$$

由瓦尔德等式我们有

$$E[N](2p-1)=i$$

或

$$E[N]=\frac{i}{2p-1}$$

7.3 回到随机游动

设

$$S_n = \sum_{i=1}^{n} X_i, \qquad n \geqslant 1$$

为随机游动。我们的第一个结果是证明，若 X_i 是只取有限个整数值的随机变量，则 $E[X]=0$ 时$\{S_n\}$是常返的。

定理 7.3.1 假设 X_i 只取值 0，±1，…，$\pm M$，$M<\infty$，则$\{S_n, n\geqslant 0\}$是常返马尔可夫链当且仅当 $E[X]=0$。

证明 $E[X]\neq 0$ 时随机游动显然是非常返的，因为它或收敛于$+\infty$(若 $E[X]>0$)，或收敛于$-\infty$(若 $E[X]<0$)。所以设 $E[X]=0$，并注意到这蕴含了$\{S_n, n\geqslant 1\}$是鞅。

以 A 记以$-M$ 到-1 的状态集合，即 $A=\{-M,-(M-1),\cdots,-1\}$。假设过程从状态 i 开始，$i\geqslant 0$。对 $j>i$，以 A_j 记状态集合 $A_j=\{j,j+1,\cdots,j+M\}$，以 N 记过程首次进入 A 或 A_j 的时刻。由定理 7.2.2，

$$E[S_N]=E[S_0]=i$$

所以

$$\begin{aligned} i &= E[S_N \mid S_N\in A]P\{S_N\in A\}+E[S_N \mid S_N\in A_j]P\{S_N\in A_j\} \\ &\geqslant -MP\{S_N\in A\}+j(1-P\{S_N\in A\}) \end{aligned}$$

或

$$P\{S_N\in A\}\geqslant \frac{j-i}{j+M}$$

因此

$$P\{\text{过程迟早进入 } A\}\geqslant P\{S_N\in A\}\geqslant \frac{j-i}{j+M}$$

令 $j\to\infty$，可见

$$P\{\text{过程迟早进入 } A \mid \text{从 } i \text{ 出发}\}=1, \qquad i\geqslant 0$$

现在令 $B=\{1,2,\cdots,M\}$。用同样的论证我们能证明对 $i\leqslant 0$

$$P\{\text{过程迟早进入 } B \mid \text{从 } i \text{ 出发}\}=1, \qquad i\leqslant 0$$

所以，对一切 i

$$P\{\text{过程迟早进入 } A\cup B \mid \text{从 } i \text{ 出发}\}=1$$

容易看出上式蕴含了将无穷多次到达有限状态集 $A\cup B$。然而，如果过程是非常返的，那么有限的状态集合只能到达有限多次。所以过程是常返的。

再设

$$S_n=\sum_{i=1}^{n}X_i,\qquad n\geqslant 1$$

为随机游动，并设 $\mu=E[X]\neq 0$。对给定的 $A,B>0$，我们将要计算 P_A，S_n 在到达小于或等于 $-B$ 的值之前达到一个大于或等于 A 的值的概率。先设 $\theta\neq 0$ 使得

$$E[e^{\theta X}]=1$$

我们将假设这样的 θ 存在(通常是唯一的)。由于

$$Z_n\equiv e^{\theta S_n}$$

是独立的均值为 1 的随机变量的乘积，故得 $\{Z_n\}$ 是均值为 1 的鞅，定义停时 N 为

$$N=\min\{n:S_n\geqslant A \text{ 或 } S_n\leqslant -B\}$$

由于能够证明定理 7.2.2 的条件(iii)被满足，我们有

$$E[e^{\theta S_N}]=1$$

所以

$$1=E[e^{\theta S_N}\mid S_N\geqslant A]P_A+E[e^{\theta S_N}\mid S_N\leqslant -B](1-P_A)\tag{7.3.1}$$

利用等式(7.3.1)能得到 P_A 的近似值如下。如果我们忽略过头的数量(越过 A 或 $-B$ 的部分)，我们有下列近似式：

$$E[e^{\theta S_N}\mid S_N\geqslant A]\approx e^{\theta A}$$

$$E[e^{\theta S_N}\mid S_N\leqslant -B]\approx e^{-\theta B}$$

从而，由(7.3.1)得

$$1\approx e^{\theta A}P_A+e^{-\theta B}(1-P_A)$$

或

$$P_A\approx\frac{1-e^{-\theta B}}{e^{\theta A}-e^{-\theta B}}\tag{7.3.2}$$

我们也能利用瓦尔德等式，然后忽略过头的数量得 $E[N]$ 的近似式，即

$$E[S_N]=E[S_N\mid S_N\geqslant A]P_A+E[S_N\mid S_N\leqslant -B](1-P_A)$$

利用近似式

$$E[S_N|S_N\geqslant A]\approx A$$
$$E[S_N|S_N\leqslant -B]\approx -B$$

我们有

$$E[S_N]\approx AP_A-B(1-P_A)$$

又由于

$$E[S_N]=E[N]E[X]$$

可见

$$E[N]\approx\frac{AP_A-B(1-P_A)}{E[X]}$$

利用 P_A 的近似式(7.3.2)得

(7.3.3) $$E[N]\approx\frac{A(1-e^{-\theta B})-B(e^{\theta A}-1)}{(e^{\theta A}-e^{-\theta B})E[X]}$$

例 7.3(a) 赌徒输光问题。设

$$X_i=\begin{cases}1, & 以概率\ p\\ -1, & 以概率\ q=1-p\end{cases}$$

我们把证明 $E[(q/p)^X]=1$ 留作习题,所以 $e^\theta=q/p$。若假设 A 与 B 是整数,那么就不会有越过头的部份,因此近似式(7.3.2)与(7.3.3)成为精确式。所以

$$P_A=\frac{1-(q/p)^{-B}}{(q/p)^A-(q/p)^{-B}}=\frac{(q/p)^B-1}{(q/p)^{A+B}-1}$$

与

$$E[N]=\frac{A(1-(q/p)^{-B})-B((q/p)^A-1)}{((p/q)^A-(p/q)^{-B})(2p-1)}$$

现在假设 $E[X]<0$,我们对随机游动迟早越过 A* 的概率感兴趣。我们要利用迄今所得结果并令 $B\to\infty$ 来计算这个概率。等式(7.3.1)说

(7.3.4) $$1=E[e^{\theta S_N}|S_N\geqslant A]P\{过程越过\ A\ 在越过\ -B\ 之前\}$$
$$+E[e^{\theta S_N}|S_N\leqslant -B]P\{过程越过\ -B\ 在越过\ A\ 之前\}$$

* 越过 A 意味着或击中 A 或超过 A。

$\theta \neq 0$ 定义为使 $E[e^{\theta x}]=1$ 成立。由于 $E[X]<0$，能够证明 $\theta>0$（见习题 13）。从而由（7.3.4）我们有

$$1 \geqslant e^{\theta A} P\{\text{过程越过 } A \text{ 在越过} -B \text{ 之前}\}$$

令 $B \to \infty$ 即得

(7.3.5) $$P\{\text{随机游动迟早越过 } A\} \leqslant e^{-\theta A}$$

7.4 应用于 G/G/1 排队系统与破产问题

7.4.1 G/G/1 排队系统

对于 $G/G/1$ 排队系统，一个顾客在排队中的等待时间的极限分布由（7.1.5）给出

(7.4.1) $$P\{D_\infty \geqslant A\} = P\{\text{对某个 } n, S_n \geqslant A\}$$

其中

$$S_n = \sum_{i=1}^{n} U_i, \qquad U_i = Y_i - X_{i+1}$$

Y_i 是第 i 个顾客的服务时间，X_{i+1} 是第 i 个与第 $i+1$ 个来到 i 之间的间隔时间。

因此，当 $E[U]=E[Y]-E[X]<0$ 时，令 $\theta>0$ 使得

$$E[e^{\theta U}] = E[e^{\theta(Y-X)}] = 1$$

由（7.3.5）有

(7.4.2) $$P\{D_\infty \geqslant A\} \leqslant e^{-\theta A}$$

存在一种情形，那时我们能得到 D_∞ 的精确分布，这就是服务时间分布是指数分布的情形。

所以假设 Y_i 是参数为 μ 的指数变量。我们记得 N 定义为 S_n 越过 A 或 $-B$ 所花的时间，在等式（7.3.4）中证明了

(7.4.3) $$\begin{aligned} 1 = & E[e^{\theta S_N} \mid S_N \geqslant A] P\{S_n \text{ 越过 } A \text{ 在越过} -B \text{ 之前}\} \\ & + E[e^{\theta S_N} \mid S_N \leqslant -B] P\{S_n \text{ 越过} -B \text{ 在越过 } A \text{ 之前}\} \end{aligned}$$

现在，$S_n = \sum_{i=1}^{n}(Y_i - X_{i+1})$，考虑 S_N 在给定 $S_N \geqslant A$ 时的条件分布，

这就是

(7.4.4) $\sum_{i=1}^{N}(Y_i-X_{i+1})$在给定$\sum_{i=1}^{N}(Y_i-X_{i+1})>A$时

的条件分布。

对N的值(例如$N=n$)及

$$X_{n+1}-\sum_{i=1}^{m-1}(Y_i-X_{i+1})$$

的值(譬如它等于c)取条件时,注意到(7.4.4)给出的条件分布正是

$$Y_n-c \qquad \text{在给定} Y_n-c>A \text{时}$$

的条件分布。由指数变量的无记忆性可得,Y在给定$Y>c+A$时的条件分布恰是$c+A$加上一个参数为μ的指数变量的分布。因此,Y_n-c在给定$Y_n-c>A$时的条件分布恰是A加上一个参数为μ的指数变量的分布。由于这对一切n及c成立,可见

$$\begin{aligned}E[e^{\theta S_N}\mid S_N\geqslant A] &= E[e^{\theta(A+Y)}]\\ &= e^{\theta A}\int e^{\theta y}\mu e^{-\mu y}dy\\ &= \frac{\mu}{\mu-\theta}e^{\theta A}\end{aligned}$$

从而,由(7.4.3)

$$\begin{aligned}1=&\frac{\mu}{\mu-\theta}e^{\theta A}P\{S_n \text{越过} A \text{在越过} -B \text{之前}\}\\ &+E[e^{\theta S_N}\mid S_N\leqslant -B]P\{S_n \text{越过} -B \text{在越过} A \text{之前}\}\end{aligned}$$

由此$\theta>0$,令$B\to\infty$得

$$1=\frac{\mu}{\mu-\theta}e^{\theta A}P\{S_n \text{迟早越过} A\}$$

故由(7.4.1)

$$P\{D_\infty\geqslant A\}=\frac{\mu-\theta}{\mu}e^{-\theta A}, \qquad A>0$$

综上所述,我们证明了下列

定理 7.4.1 对具有 iid 服务时间$Y_i, i\geqslant 1$,及 iid 来到间隔时间$X_1,X_2,\cdots$的$G/G/1$排队系统,$E[Y]<E[X]$时,

$$P\{D_\infty \geqslant A\} \leqslant e^{-\theta A}$$

其中 $\theta > 0$ 使得

$$E[e^{\theta Y}]E[e^{-\theta X}] = 1$$

此外，若 Y_i 是参数为 μ 的指数分布，则

$$P\{D_\infty \geqslant A\} = \frac{\mu - \theta}{\mu} e^{-\theta A}, \quad A > 0$$

$$P\{D_\infty = 0\} = \frac{\theta}{\mu}$$

这时的 θ 使得

$$E[e^{-\theta X}] = \frac{\mu - \theta}{\mu}$$

7.4.2 一个破产问题

假设一家保险公司收到的赔偿要求遵循一个更新过程，其来到间隔时间为 $X_1, X_2, \cdots$。假设相继的要求赔偿数是 iid，且独立于它们发生的更新过程。以 Y_i 记第 i 个要求赔偿数。因此若以 $N(t)$ 记到 t 为止要求赔偿的次数，则保险公司到 t 为止的赔偿总数是 $\sum_{i=1}^{N(t)} Y_i$。另一方面，假设公司以每单位时间为 c 的常数速率收到保险费，$c > 0$。我们感兴趣的是确定保险公司从初始资本 A 开始最终破产的概率。也就是要求

$$p = P\{\text{对某个 } t \geqslant 0, \sum_{i=1}^{N(t)} Y_i > ct + A\}$$

显然，若 $E[Y] \geqslant cE[X]$，以概率 1 公司最终要破产(为什么?)。所以我们将假定

$$E[Y] < cE[X]$$

如果公司破产，那么这事件将发生在有赔偿要求发生的时刻，这一点也是相当显然的(因为只有在赔偿要求发生的时刻公司的资产才减少)。在第 n 次要求发生之前的时刻，公司的财产为

$$A + c\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i$$

因此我们想要求的概率，记为 $p(A)$，是

$$p(A)=P\{\text{对某个 } n, A+c\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}Y_i<0\}$$

或等价地

$$p(A)=P\{\text{对某个 } n, S_n>A\}$$

其中

$$S_n=\sum_{i=1}^{n}(Y_i-cX_i)$$

是随机游动。由(7.4.1)可见

(7.4.5) $$p(A)=P\{D_\infty>A\}$$

其中 D_∞ 是一个来到间隔时间为 cX_i，服务时间为 Y_i 的 $G/G/1$ 排队系统的在排队中的极限等待时间。因此从定理 7.4.1 我们有下列

定理 7.4.2 (i)保险公司迟早破产的概率，记为 $p(A)$，满足

$$p(A)\leqslant e^{-\theta A}$$

其中 θ 使得

$$E[e^{\theta(Y_i-cX_i)}]=1$$

(ii)若要求赔偿数服从参数为 μ 的指数分布，则

$$p(A)=\frac{\mu-\theta}{\mu}e^{-\theta A}$$

其中 θ 使得

$$E[e^{-\theta cX}]=\frac{\mu-\theta}{\mu}$$

(iii)若赔偿要求来到的过程是参数为 λ 的泊松过程，则

$$p(0)=\lambda E[Y]/c$$

证明 (i)及(ii)直接来自定理 7.4.1。在(iii)中，cX 是参数为 λ/c 的指数变量，从而由(7.4.5)$p(0)$等于 $M/G/1$ 排队系统中顾客在排队中的极限等待时间为正的概率。但这正是在 $M/G/1$ 排队系统中一个来到的顾客发现系统不空的概率。因为它是一个有泊松来到的系统，一个来客所看到的状态的极限分布等同于在时刻 t 的系统状态的极限分布。(这是因为由于泊松来到的假定，在一个顾客于时刻 t 来到的条件下，时刻 t 的系统状态的分布等同

于时刻 t 的状态的无条件分布。)因此一个来客发现系统不空的(极限)概率等于系统不空的极限概率,正如我们用多种不同的方法证明的,它等于来到速率乘上平均服务时间(见第四章的例 4.3(a))。

7.5 直线上的布莱克威尔定理

设$\{S_n,n\geqslant 1\}$为随机游动,$0<\mu=E[X]<\infty$。以 $U(t)$记使 $S_n\leqslant t$ 的 n 的个数。也就是

$$U(t)=\sum_{n=1}^{\infty}I_n, \qquad 其中\ I_n=\begin{cases}1, & 若\ S_n\leqslant t\\ 0, & 其它\end{cases}$$

若 X_i 是非负的,则 $U(t)$正是 $N(t)$,到 t 为止的更新数。

令 $u(t)=E[U(t)]$。本节我们将证明一个类似于布莱克威尔定理的结果。

布莱克威尔定理 若 $\mu>0$,X_i 不是格点的,则对 $a>0$,$t\to\infty$ 时

$$u(t+a)-u(t)\to a/\mu$$

在给出上述定理的证明之前,介绍上升及下降阶梯变量的概念是有用的。若

$$S_n>\max(S_0,S_1,\cdots,S_{n-1}), \quad 其中\ S_0\equiv 0$$

我们说在时刻 n 发生一个阶梯高度为 S_n 的上升阶梯变量。也就是每当随机变量达到一个新的高度时发生一个上升阶梯变量。例如,随机游动首次为正的时刻发生第一个上升阶梯变量。如果在时刻 n 发生一个高度为 S_n 的阶梯变量,那么下一个阶梯变量将在第一个使

$$S_{n+j}>S_n$$

的时刻 $n+j$ 发生,或等价地在第一个使

$$X_{n+1}+\cdots+X_{n+j}>0$$

的时刻 $n+j$ 发生。由于诸 X_i 是独立同分布的,因此随机游动在阶梯变量间的变化彼此都是概率意义上的复制品。也就是,若以 N_i

记第 $i-1$ 个与第 i 个阶梯变量之间的时间，则随机向量$(N_i, S_{N_i}-S_{N_{i-1}})(i\geqslant 1)$是独立同分布的（其中 $S_{N_0}\equiv 0$）。

类似地我们能定义下降阶梯变量的概念，当随机游动达到一个新的低度时就说它们发生。以 $p(p_*)$记迟早得到一个上升（下降）阶梯变量的概率。也就是

$$p=P\{对某个 n, S_n>0\}$$

$$p_*=P\{对某个 n, S_n<0\}$$

每当得到一个上升（下降）阶梯变量之后，就又以同样的概率 $p(p_*)$再得到另一个。因此恰有 n 个这样的上升（下降）阶梯变量$(n\geqslant 0)$的概率为 $p^n(1-p)(p_*^n(1-p_*))$。所以上升（下降）阶梯变量的个数为有限且均值有限当且仅当 $p(p_*)$小于 1。由于 $E[X]>0$，由强大数定律可得，以概率 1，$n\to\infty$时 $S_n\to\infty$；所以，以概率 1 有无穷多个上升阶梯变量，但只有有限个下降阶梯变量。因此，$p=1$ 及 $p_*<1$。

现在我们已准备好去证明布莱克威尔定理。证明分几步走。先论证 $t\to\infty$时 $u(t+a)-u(t)$趋于一个极限。然后证明这个极限值等于一个常数乘以 a；最后证明基本更新定理的一个推广，使我们能把这个常数等同于 $1/\mu$。

布莱克威尔定理的证明 相继的上升阶梯高度构成一个更新过程。以 $Y(t)$记这个更新过程在时刻 t 的剩余寿命。也就是，$t+Y(t)$是随机游动第一个超过 t 的值。现在易见，给定 $Y(t)$的值，譬如 $Y(t)=y$ 时 $U(t+a)-U(t)$的分布不依赖于 t。也就是，如果我们知道随机过程第一个超过 t 的点发生于过 t 距离 y 处，则$(t,t+a)$中点的个数与已知第一个正值是 y 时$(0,a)$中点的个数同分布。因此，对某个函数 g，

$$E[U(t+a)-U(t)|Y(t)]=g(Y(t))$$

取期望故得

$$u(t+a)-u(t)=E[g(Y(t))]$$

$Y(t)$，作为一个非格点更新过程在 t 时的剩余寿命，收敛于一个极限分布（即来到间隔分布的平衡分布）。因此，$E[g(Y(t))]$将收敛于 $E[g(Y(\infty))]$，其中 $Y(\infty)$服从 $Y(t)$的极限分布，从而我们证明了

$$\lim_{t\to\infty}[u(t+a)-u(t)]$$

的存在性。现在令

$$h(a)=\lim_{t\to\infty}[u(t+a)-u(t)]$$

则有

$$\begin{aligned}h(a+b)&=\lim_{t\to\infty}[u(t+a+b)-u(t+b)+u(t+b)-u(t)]\\&=\lim_{t\to\infty}[u(t+b+a)-u(t+b)]+\lim_{t\to\infty}[u(t+b)-u(t)]\\&=h(a)+h(b)\end{aligned}$$

这蕴含了,对某个常数 c

(7.5.1) $$h(a)=\lim_{t\to\infty}[u(t+a)-u(t)]=ca$$

为辨认 c 的值,以 N_t 记使 $S_n>t$ 的第一个 n。若 X_i 是有界的,譬如说以 M 为界,则

$$t<\sum_{i=1}^{N_t}X_i\leqslant t+M$$

取期望并利用瓦尔德等式(使用类似于命题 7.1.1 中用的论证方法可得 $E[N_t]<\infty$)得

$$t<E[N_t]\mu\leqslant t+M$$

所以 $t\to\infty$ 时

(7.5.2) $$\frac{E[N_t]}{t}\to\frac{1}{\mu}$$

若 X_i 不是有界的,则我们能用截尾证法(正如在证基本更新定理时所用的那样)证明(7.5.2)成立。现在 $U(t)$ 能表示为

(7.5.3) $$U(t)=N_t-1+N_t^*$$

其中 N_t^* 是走过 t 之后 S_n 落入$[-\infty,t]$的次数。由于随机变量 N_t^* 不大于时刻 N_t 之后使随机游动小于 S_{N_t} 的点的个数,故得

(7.5.4) $$E[N_t^*]\leqslant E[\text{使 } S_n<0 \text{ 的 } n \text{ 的个数}]$$

现在我们将论证上式的右边是有限的,故由(7.5.2)及(7.5.3)得 $t\to\infty$ 时

(7.5.5) $$\frac{u(t)}{t}\to\frac{1}{\mu}$$

(7.5.4)的右边为有限的论证如下:由命题 7.1.1 注意到,当 N 是第一个使 $S_n>0$ 的 n 的值时 $E[N]<\infty$。在时刻 N 以一个正概率 $1-p^*$ 随机游动的将来的值永不低于 S_N。若一个将来的值确实低于 S_N,那么再用类似于命题 7.1.1 中用过的证法可知,直到随机游动重新变为正为止的平均附加时间是

有限的。那时又以正概率 $1-p^*$ 将来的值永不低于现在的正值，等等。我们能以此作为证明

$$E[\text{使 } S_n<0 \text{ 的 } n \text{ 的个数}]\leqslant \frac{E[N\mid X_1<0]}{1-p^*}<\infty$$

的基础。这就说明了(7.5.5)。

现在将利用(7.5.1)及(7.5.5)完成证明。从(7.5.1)我们有 $i\to\infty$ 时

$$u(i+1)-u(i)\to c$$

这蕴含 $n\to\infty$ 时

$$\sum_{i=1}^{n}\frac{u(i+1)-u(i)}{n}\to c$$

或等价地

$$\frac{u(n+1)-u(1)}{n}\to c$$

由(7.5.5)这蕴含了 $c=1/\mu$。定理证毕。

注记　上述证明有一处不严格，即尽管 $Y(t)$ 的分布收敛于 $Y(\infty)$ 的分布，不一定能得出 $E[g(Y(t))]$ 的收敛于 $E[g(Y(\infty))]$。我们应该直接证明其收敛性。

7.6 再论鞅

对一切 n，$E[|Z_n|]<\infty$ 的随机过程 $\{Z_n,n\geqslant 1\}$ 称为下鞅，若

(7.6.1)
$$E[Z_{n+1}\mid Z_1,\cdots,Z_n]\geqslant Z_n$$

称为上鞅，若

(7.6.2)
$$E[Z_{n+1}\mid Z_1,\cdots,Z_n]\leqslant Z_n$$

因此一个下鞅是一种上偏的赌博，一个上鞅是下偏的赌博。

从(7.6.1)可见，对下鞅

$$E[Z_{n+1}]\geqslant E[Z_n]$$

对上鞅这不等式反向。类似于定理 7.2.2 的结论对下鞅与上鞅仍然成立。特别，若 N 是 $\{Z_n,n\geqslant 1\}$ 的停时，使得定理 7.2.2 中的充分条件有一个被满足，则对下鞅

$$E[Z_N]\geqslant E[Z_1]$$

对上鞅这不等式反向。

最重要的鞅定理是收敛定理。在讨论这个结果之前我们需要一些准备知识。首先回忆下列凸函数的定义。

定义 实轴上的函数 f 称为凸的，若对一切 $x_1, x_2, 0<\lambda<1$，

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda)f(x_2)$$

我们将需要下面的引理。

引理 7.6.1 延森不等式

若 f 是凸函数，则

$$E[f(X)] \geqslant f(E[X])$$

假定期望都存在。

引理 7.6.2 马尔可夫不等式

若 Y 是非负随机变量，则对一切 $a>0$

$$P\{Y \geqslant a\} \leqslant E[Y]/a$$

引理 7.6.1 与 7.6.2 的证明留作习题。

引理 7.6.3

若 $\{Z_i, i\geqslant 1\}$ 是下鞅，N 是停时使得 $P\{N\leqslant n\}=1$，则

$$E[Z_1] \leqslant E[Z_N] \leqslant E[Z_n]$$

证明 由于 N 有界，以类似于定理 7.2.2 的结论可知 $E[Z_N]\geqslant E[Z_1]$。现在有

$$\begin{aligned} E[Z_n | Z_1, \cdots, Z_N, N = k] &= E[Z_n | Z_1, \cdots, Z_k, N = k] \\ &= E[Z_n | Z_1, \cdots, Z_k] \quad (\text{为什么?}) \\ &\geqslant Z_k \\ &= Z_N \end{aligned}$$

对上式取期望即得结论。

引理 7.6.4

若 $\{Z_n, n\geqslant 1\}$ 是鞅，f 是凸函数，则 $\{f(Z_n), n\geqslant 1\}$ 是下鞅。

证明

$$E[f(Z_{n+1}) | Z_1, \cdots, Z_n] \geqslant f(E[Z_{n+1} | Z_1, \cdots, Z_n]) \quad (\text{由延森不等式})$$

$$=f(Z_n)$$

定理 7.6.5(下鞅的柯尔莫哥洛夫不等式)　若$\{Z_n,n\geqslant 1\}$是非负下鞅,则对$a>0$,

$$P\{\max(Z_1,\cdots,Z_n)>a\}\leqslant\frac{E[Z_n]}{a}$$

证明　以N记使$i\leqslant n,Z_i>a$的最小的i;若对一切$i=1,\cdots,n,Z_i\leqslant a$,定义它等于$n$。注意到$\max(Z_1,\cdots,Z_n)>a$等价于$Z_N>a$。所以

$$\begin{aligned}P\{\max(Z_1,\cdots,Z_n)>a\}&=P\{Z_N>a\}\\&\leqslant E[Z_N]/a(\text{由马尔可夫不等式})\\&\leqslant E[Z_n]/a\end{aligned}$$

其中最后的不等式得自引理 7.6.3,因为$N\leqslant n$。

系 7.6.6

设$\{Z_n,n\geqslant 1\}$是鞅,则对$a>0$

(i) $P\{\max(|Z_1|,\cdots,|Z_n|)>a\}\leqslant E[|Z_n|]/a$

(ii) $P\{\max(|Z_1|,\cdots,|Z_n|)>a\}\leqslant E[|Z_n^2|]/a^2$

证明　(i)及(ii)得自引理 7.6.4 与柯尔莫哥洛夫不等式,因为函数$f(x)=|x|$与$f(x)=x^2$都是凸的。

现在我们正为鞅收敛定理作好准备。

定理 7.6.7(鞅收敛定理)　设$\{Z_n,n\geqslant 1\}$为鞅,使得对某个$M<\infty$

$$E[|Z_n|]\leqslant M$$

则以概率 1,$\lim_{n\to\infty}Z_n$存在且有限。

证明　我们将在较强的假设:$E[Z_n^2]$有界之下证明定理(较强是因为$E[|Z_n|]\leqslant(E[Z_n^2])^{1/2}$)。因为$f(x)=x^2$是凸的,由引理 7.6.4 得出。$\{Z_n^2,n\geqslant 1\}$是下鞅;从而$E[Z_n^2]$是非减的。由于$E[Z_n^2]$有界,故$n\to\infty$时它收敛。令$\mu<\infty$为极限:

$$\mu=\lim_{n\to\infty}E[Z_n^2]$$

我们通过证明$\{Z_n,n\geqslant 1\}$以概率 1 是柯西序列而证得$\lim_{n\to\infty}Z_n$存在且有限,也就是将证明以概率 1,k、$m\to\infty$时

$$|Z_{m+k}-Z_m|\to 0$$

现在

(7.6.3) $$P\{|Z_{m+k}-Z_m|>\varepsilon, \text{对某个 } k\leqslant n\}$$
$$\leqslant E[(Z_{m+n}-Z_m)^2]/\varepsilon^2 \qquad \text{（由柯尔莫哥洛夫不等式）}$$
$$=E[Z_{m+n}^2-2Z_mZ_{m+n}+Z_m^2]/\varepsilon^2$$

但是

$$E[Z_mZ_{m+n}] = E[E[Z_mZ_{m+n}|Z_m]]$$
$$=E[Z_mE[Z_{m+n}|Z_m]]$$
$$=E[Z_m^2]$$

从而由(7.6.3)

$$P\{|Z_{m+k}-Z_m|>\varepsilon, \text{对某个 } k\leqslant n\}\leqslant \frac{E[Z_{m+n}^2]-E[Z_m^2]}{\varepsilon^2}$$

令 $n\to\infty$，回忆 μ 的定义得

$$P\{|Z_{m+k}-Z_m|>\varepsilon, \text{对某个 } k\}\leqslant \frac{\mu-E[Z_m^2]}{\varepsilon^2}$$

所以，$m\to\infty$ 时

$$P\{|Z_{m+k}-Z_m|>\varepsilon, \text{对某个 } k\}\to 0$$

从而，以概率 1，Z_n 是柯西序列，故 $\lim_{n\to\infty}Z_n$ 存在且有限。

系 7.6.8

若 $\{Z_n, n\geqslant 0\}$ 是非负鞅，则以概率 1，$\lim_{n\to\infty}Z_n$ 存在且有限。

证明 由于 Z_n 非负

$$E[|Z_n|]=E[Z_n]=E[Z_1]$$

例 7.6(a) 分支过程。若 X_n 是分支过程中第 n 代的个体总量，每个个体的平均后代数为 m，则 $Z_n=X_n/m^n$ 是非负鞅。从而由系 7.6.8，$n\to\infty$ 时它将收敛。由此我们能得出结论：或 $X_n\to 0$，或它以指数速率趋于 ∞。

例 7.6(b) 一个关于赌博的结果。考虑一个赌徒参加公平的赌博，即若 Z_n 是赌徒在 n 局之后的赌金，则 $\{Z_n, n\geqslant 1\}$ 是鞅。现在假设不给赊钱，所以赌徒的赌金不允许变成负数，并设每一局至少赢或输 1 元。以

$$N=\min\{n: Z_n=Z_{n+1}\}$$

记直到赌徒被强迫退出为止已玩的赌局数。（因为 $Z_n-Z_{n+1}=0$，他没有赌第 $n+1$ 局。）由于 $\{Z_n\}$ 是非负鞅，由收敛定理可见

(7.6.4) $$\lim_{n\to\infty}Z_n \text{ 以概率 1 存在且有限。}$$

但对 $n<N$，$|Z_{n+1}-Z_n|\geqslant 1$，所以(7.6.4)蕴含了，以概率 1

$$N<\infty$$

也就是，以概率 1 赌徒最终要输光。

我们将以利用鞅收敛定理证明强大数定律结束本章。

定理 7.6.9(强大数定律)　设 $X_1,X_2,\cdots$ 独立同分布，均值 μ 有限，$S_n=\sum_{i=1}^{n}X_i$，则

$$P\{\lim_{n\to\infty}S_n/n=\mu\}=1$$

证明　我们将在矩母函数 $\Psi(t)=E[e^{tX}]$ 存在的假设下证明定理。

对给定的 $\varepsilon>0$，定义 $g(t)$ 为

$$g(t)=e^{t(\mu+\varepsilon)}/\Psi(t)$$

由于

$$g(0)=1,$$

$$g'(0)=\frac{\Psi(t)(\mu+\varepsilon)e^{t(\mu+\varepsilon)}-\Psi'(t)e^{t(\mu+\varepsilon)}}{\Psi^2(t)}\bigg|_{t=0}=\varepsilon>0$$

存在一个值 $t_0>0$ 使得 $g(t_0)>1$。现在我们证明 S_n/n 只能有限多次与 $\mu+\varepsilon$ 一样大。注意到

$$(7.6.5)\qquad \frac{S_n}{n}\geqslant\mu+\varepsilon\quad\Rightarrow\quad\frac{e^{t_0S_n}}{\Psi^n(t_0)}\geqslant\left(\frac{e^{t_0(\mu+\varepsilon)}}{\Psi(t_0)}\right)^n=(g(t_0))^n$$

然而，$e^{t_0S_n}\Psi^n(t_0)$ 是独立的均值为 1 的随机变量(第 i 个变量是 $e^{t_0X_i}/\Psi(t)$)的乘积，故而是鞅。由于它也是非负数，收敛定理证明了，以概率 1

$$\lim_{n\to\infty}e^{t_0S_n}/\Psi^n(t_0)\ \text{存在且有限。}$$

由于 $g(t_0)>1$，由(7.6.5)得出

$$P\{\text{对无穷多个}\,n,\frac{S_n}{n}>\mu+\varepsilon\}=0$$

类似地，定义函数 $f(t)=e^{t(\mu-\varepsilon)}/\Psi(t)$，注意到 $f(0)=1$，$f'(0)=-\varepsilon$，存在一个值 $t_0<0$ 使 $f(t_0)>1$，以同样的方式我们能证明

$$P\{\text{对无穷多个}\,n,\frac{S_n}{n}\leqslant\mu-\varepsilon\}=0$$

因此

$$P\{\text{对一切}\,n,\text{除了有限个},\mu-\varepsilon\leqslant\frac{S_n}{n}\leqslant\mu+\varepsilon\}=1$$

由于上式对一切 $\varepsilon>0$ 成立，故

$$P\{\lim_{n\to\infty}\frac{S_n}{n}=\mu\}=1$$

习 题

1. 考虑下述水流进与流出水库的模型。假设在第 n 天，y_n 单位的水量从外界水源，如降雨与河流，流入水库。在每一天结束时按下述规则放水：如果库内水量大于 a，放掉水量 a，若它小于或等于 a，则放掉全部水量。水库的容量为 C，且一旦达到容量，多余的要流进水库的水假设全部流失。例如，在第 n 天开始时水量为 x，则在这天结束时（放水之前）水量是 $\min(x+y_n, C)$。以 S_n 记第 n 天结束时放水之后的水库水量。假设 $y_n(n\geqslant 1)$ 独立同分布，证明：$\{S_n, n\geqslant 1\}$ 是一个随机游动，在 0 及 $C-a$ 处有反射壁。
2. 对简单随机游动计算到达状态 k 的平均次数。
3. 设 $X_1, X_2, \cdots, X_n$ 可交换。计算 $E[X_1|X_{(1)}, X_{(2)}, \cdots, X_{(n)}]$，其中 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 是 X_i 的顺序统计量。
4. 若 $X_1, X_2, \cdots$ 是可交换随机变量的无穷序列，$E[X_1^2]<\infty$，证明 $\mathrm{Cov}(X_1, X_2)\geqslant 0$。（提示：注意 $\mathrm{Var}(\sum_{i=1}^{n}X_i)$。）当可交换随机变量集合有限时给出一个反例。
5. 对鞅 $\{Z_n, n\geqslant 1\}$，令 $X_i=Z_i-Z_{i-1}$，其中 $Z_0\equiv 0$。证明：
$$\mathrm{Var}(Z_n)=\sum_{i=1}^{n}\mathrm{Var}(X_i)$$
6. 验证：当 X_n 是一个分支过程的第 n 代的总量，每个个体的平均后代数是 m 时，$\{X_n/m^n, n\geqslant 1\}$ 是鞅。
7. 对简单随机游动，$p\neq\frac{1}{2}$，论证 $(q/p)^{S_n}(n\geqslant 1)$ 是鞅。
8. 若
$$S_n=\sum_{i=1}^{n}X_i,\quad n\geqslant 1$$
是随机游动，$E[X_i]=0$ 及 $\mathrm{Var}(X_i)=\sigma^2$，证明：$Z_n=S_n^2-n\sigma^2$ 时，$\{Z_n, n\geqslant 1\}$ 是鞅。
9. 设 Y 为均值有限的随机变量，$W_1, W_2, \cdots$ 为随机向量。证明：$Z_n=E[Y|W_1, \cdots, W_n]$ 时，$\{Z_n, n\geqslant 1\}$ 是鞅。考虑一个人玩一种他最终要赢或输

的赌博。以 P_n 记已知前 n 局赌博的结果时他最终要赢的条件概率。论证 P_n 是鞅。

10. 连续地掷一枚硬币，正面出现的概率为 p。求直到下列花样出现为止的平均投掷次数：

(a) HHTTHH

(b) HTHTT　　(H—正面，T—反面)

现在假设硬币是均匀的，即 $p=\frac{1}{2}$；以 A 及 B 记花样 $A=$HTHT 及 $B=$THTT。求 $E[N_A]$，$E[N_B]$及 $P\{A$ 在 B 之前出现$\}$。

11. 论证 X_i 只取值 $0,\pm 1,\cdots,\pm M$，且 $E[X_i]=0$ 的随机游动是零常返的。

12. 以 $S_n(n\geqslant 0)$ 记一个随机游动，且

$$\mu=E[S_{n+1}-S_n]\neq 0$$

对 $A>0,B>0$ 令

$$N=\min\{n: S_n\geqslant A \text{ 或 } S_n\leqslant -B\}$$

证明：$E[N]<\infty$。(提示：论证存在一个 k 使得 $P\{S_k>A+B\}>0$。然后证明 $E[N]\leqslant kE[G]$，其中 G 是一个适当定义的几何分布随机变量。)

13. 利用延森不等式(f 为凸函数时，$E[f(X)\geqslant f(E[X])]$)，证明：若 $\theta\neq 0$，$E[X]<0$，$E[e^{\theta X}]=1$，则 $\theta>0$。

14. 在 7.4 节中的保险破产问题中解释为什么若 $E[Y]\geqslant cE[X]$，公司以概率 1 最终破产。

15. 在 7.4 节中以 F 记赔偿要求来到间隔的分布，以 G 记赔偿数的分布。证明：从资产 A 开始的公司迟早破产的概率 $p(A)$满足

$$p(A)=\int_0^\infty\int_0^{A+ct}p(A+ct-x)dG(x)dF(t)+\int_0^\infty\overline{G}(A+ct)dF(t)$$

16. 对一个 $\mu=E[X]>0$ 的随机游动，论证以概率 1，$t\to\infty$时

$$\frac{u(t)}{t}\to\frac{1}{\mu}$$

其中 $u(t)$等于使 $0\leqslant S_n\leqslant t$ 的 n 的个数。

17. 设$S_n=\sum_{i=1}^{n}X_i$为一个随机游动，以 λ_i，$i>0$，记一个阶梯高度等于 i 的概率，即 $\lambda_i=P\{S_n$ 的第一个正值等于 $i\}$。

(a) 证明：若

$$P\{X_i=j\}=\begin{cases}q, & j=-1,\\ \alpha_j, & j\geqslant 1,\end{cases}\qquad q+\sum_{j=1}^{\infty}\alpha_j=1$$

则 λ_i 满足

$$\lambda_i = \alpha_i + q(\lambda_{i+1} + \lambda_1\lambda_i), \quad i > 0$$

(b) 若 $P\{X_i = j\} = \frac{1}{5}, j = -2, -1, 0, 1, 2$，证明：

$$\lambda_1 = \frac{1+\sqrt{5}}{3+\sqrt{5}}, \qquad \lambda_2 = \frac{2}{3+\sqrt{5}}$$

18. 设 $S_n(n \geqslant 0)$ 为随机游动，其中 X_i 有分布 F。以 $G(t,s)$ 记第一个超过 t 的 S_n 的值小于或等于 $t+s$ 的概率，即

$$G(t,s) = P\{\text{第一个超过}\, t\, \text{的和} \leqslant t + s\}$$

证明：

$$G(t,s) = F(t+s) - F(t) + \int_{-\infty}^{t} G(t-y,s)dF(y)$$

19. 证明延森不等式。

20. 证明马尔可夫不等式。

21. 一个罐子最初装有一个白球及一个黑球。每次从中取出一球，放回时再放进一个同样颜色的球。以 Z_n 记第 n 次放回之后罐中白球的比例。

(a) 证明：$\{Z_n, n \geqslant 1\}$ 是鞅。

(b) 证明：罐中白球的比例大到 $\frac{3}{4}$ 的概率至多为 $\frac{2}{3}$。

22. 若 f 是递增凸函数，$\{Z_n, n \geqslant 0\}$ 是下鞅，证明：$\{f(Z_n), n \geqslant 0\}$ 也是下鞅。

23. 设 $X_1, X_2, \cdots$ 相互独立，$E[X_i] = \mu_i$，$\mathrm{Var}(X_i) = \sigma_i^2$，$i \geqslant 1$。证明：

$$P\left\{\max_{k=1,\cdots,n}\left|\sum_{i=1}^{k}(X_i - \mu_i)\right| > a\right\} \leqslant \sum_{i=1}^{n}\frac{\sigma_i^2}{a^2}$$

这称为柯尔莫哥洛夫不等式。

24. 考虑一系列独立地投郑硬币，以 $P\{\text{正面}\}$ 记每次投掷中出现正面的概率。以 A 记假设 $P\{\text{正面}\} = a$，以 B 记假设 $P\{\text{正面}\} = b$，$0 < a, b < 1$。以 X_i 记第 i 次投掷的结果，令

$$Z_n = \frac{P\{X_1, \cdots, X_n | A\}}{P\{X_1, \cdots, X_n | B\}}$$

证明：若假设 B 为真，则

(a) $\{Z_n\}$ 是鞅；

(b) 以概率 1，$\lim_{n\to\infty} Z_n$ 存在；

(c) 若 $b \neq a$，$\lim_{n\to\infty} Z_n$ 是什么？

25. 设 $Z_n, n \geqslant 1$，是一列随机变量，使得 $Z_1 \equiv 1$，且对 $n > 1$，给定$Z_1, \cdots, Z_{n-1}$时 Z_n 是以 Z_{n-1}为均值的泊松变量。n 大时关于 Z_n 我们能说些什么？

26. 设 $X_1, X_2, \cdots$为独立随机变量，使得

$$P\{X_i = -1\} = 1 - 1/2^i$$
$$P\{X_i = 2^i - 1\} = 1/2^i, \quad i \geqslant 1$$

利用此序列构造一个零均值鞅 Z_n，使得以概率 1，$\lim\limits_{n \to \infty} Z_n = -\infty$。（提示：利用波莱尔—坎泰利引理。）

27. 连续时间过程$\{X(t), t \geqslant 0\}$称为鞅，若对一切 $s \leqslant t$，

$$E[X(t) \mid X(u), 0 \leqslant u \leqslant s] = X(s)$$

(a) 若$\{Z(t)\}$是漂移系数为 μ 的布朗运动，证明：$X(t) = Z(t) - t\mu$ 时，$\{X(t)\}$是鞅。

(b) 若$\{N(t)\}$是强度函数为 $\alpha(t)$的非齐次泊松过程，证明：$X(t) = N(t) - \int_0^t \alpha(s) ds$ 时，$\{X(t)\}$是鞅。假设连续时间情形时与定理 7.2.2 类似的结果成立，所需的正则性条件满足，证明：

(c) 对具有漂移 $\mu, \mu > 0$ 的布朗运动，从 0 到 x 的平均时间是 $x/\mu, x > 0$；

(d) $n = E[\int_0^{T_n} \alpha(s) ds]$，其中 T_n 是一个具有强度函数 $\alpha(s)$的非齐次泊松过程击中 n 的时刻。

参考文献

关于随机游动的其它结果读者应参看文献 1，4，特别是 8。关于可交换随机变量的补充结果应参看文献 2。鞅为杜勃所发展，他的教科书（文献 3）仍然是一本标准参考书。文献 5 给出了鞅的一个很好的综述，其数学水平略高于本书。例 7.2(a)的结果取自文献 7；平均步子为零且只取有限个值的随机游动是常返的，即定理 7.3.1 的鞅证法取自文献 6。7.5 节中布莱克威尔定理的证法我们相信是新的。

1. D. R. Cox and H. D. Miller, *Theory of Stochastic Processes*, Methuen, London, 1965。

2. B. Definetti, *Theory of Probability*, Vols. I and II, Wiley, New York, 1970。

3. J. Doob, *Stochastic Processes*, Wiley, New York, 1953。

4. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, Wiley, New York, 1966。
5. S. Karlin and H. Taylor, *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York, 1975。
6. J. Kemeny, L. Snell, and A. Knapp, *Denumerable Markov Chains*, Van Nostrand, New Jersey, 1966。
7. S. Y. R. Li, "A Martingale Approach to the Study of the Occurrence of Patterns in Repeated Experiments," *Annals of Probability*, 8, No. 6(1980), pp. 1171—1175。
8. F. Spitzer, *Principles of Random Walk*, Van Nostrand, New Jersey, 1964。

8 随机序关系

引 言

本章介绍一些随机变量之间的随机序关系。在 8.1 节中，我们考虑一个随机变量随机大于另一个随机变量的概念，给出了它对具有单调失效率函数的随机变量的应用。在 8.2 节中，我们继续研究随机大于的概念，介绍了耦合方法，并阐明它的用处。特别，在 8.2.1 节中我们用耦合方法建立了生灭过程的一些随机单调性质，在 8.2.2 节证明了：有限状态遍历马尔可夫链的 n 步转移概率以指数速度收敛于它的极限概率。

在 8.3 节中，我们考虑随机变量间的失效率序，它强于随机大于序。我们表明了如何利用这个思想去比较某些计数过程，事实上我们利用它证明了更新理论中来到间隔分布连续时的布莱克威尔定理。也给出了来到间隔分布有递减失效率的更新过程的一些单调性质。在 8.4 节，我们考虑似然比序。

在 8.5 节，我们考虑一个随机变量比另一个随机变量更多变的概念，且在 8.6 节中，将它用于比较(i)排队系统(8.6.1 节)，(ii)一个更新过程和一个泊松过程(8.6.2 节)，(iii)分支过程。

8.1 随机大于

我们说随机变量 X 随机地大于随机变量 Y，如果对一切 a

(8.1.1) $$P\{X>a\}\geqslant P\{Y>a\}$$

记作 $X\underset{\text{st}}{\geqslant}Y$。如果 X 和 Y 分别服从分布 F 和 G，那么(8.1.1)等价于对一切 a

$$\overline{F}(a)\geqslant\overline{G}(a)$$

引理 8.1.1

如果 $X\underset{\text{st}}{\geqslant}Y$，则 $E[X]\geqslant E[Y]$

证明 先设 X，Y 是非负随机变量，那么

$$E[X]=\int_0^\infty P\{X>a\}da\geqslant\int_0^\infty P\{Y>a\}da=E[Y]$$

一般地，我们可以把任一随机变量 Z 写成二个非负随机变量之差如下：

$$Z=Z^+-Z^-$$

其中

$$Z^+=\begin{cases}Z, & 若 Z\geqslant 0\\ 0, & 若 Z<0\end{cases}\qquad Z^-=\begin{cases}0, & 若 Z\geqslant 0\\ -Z, & 若 Z<0\end{cases}$$

现在我们把证明

$$X\underset{\text{st}}{\geqslant}Y\ \Rightarrow\ X^+\underset{\text{st}}{\geqslant}Y^+,\qquad X^-\underset{\text{st}}{\leqslant}Y^-$$

留作练习。由此得，

$$E[X]=E[X^+]-E[X^-]\geqslant E[Y^+]-E[Y^-]=E[Y]$$

下列命题给出了随机大于的另一个定义。

命题 8.1.2

$X\underset{\text{st}}{\geqslant}Y\Leftrightarrow$对一切递增函数 f，$E[f(X)]\geqslant E[f(Y)]$

证明 先设 $X\underset{\text{st}}{\geqslant}Y$，$f$ 是一个递增函数。我们证明 $f(X)\underset{\text{st}}{\geqslant}f(Y)$ 如下。令 $f^{-1}(a)=\inf\{x;f(x)>a\}$，则

$$\begin{aligned}P\{f(X)>a\}&=P\{X>f^{-1}(a)\}\geqslant P\{Y>f^{-1}(a)\}\\&=P\{f(Y)>a\}\end{aligned}$$

从而，$f(X)\underset{\text{st}}{\geqslant}f(Y)$。所以由引理 8.1.1 得 $E[f(X)]\geqslant E[f(Y)]$。

现在设对一切递增函数 f，有 $E[f(X)]\geqslant E[f(Y)]$。对任一 a，以 f_a 记递增函数：

$$f_a(x)=\begin{cases}1, & \text{若 } x>a\\ 0, & \text{若 } x\leqslant a\end{cases}$$

那么

$$E[f_a(X)]=P\{X>a\},\qquad E[f_a(Y)]=P\{Y>a\}$$

由此可见 $X\underset{\text{st}}{\geqslant}Y$。

例 8.1(a) 递增和递减失效率。 设 X 为一个具有分布函数 F 和密度函数 f 的非负随机变量。我们记得 X 的失效率函数定义为

$$\lambda(t)=f(t)/\overline{F}(t)$$

我们说 X 是一个递增失效率(IFR)随机变量，如果

$$\lambda(t)\uparrow t\quad（关于\ t\ 递增）$$

我们说 X 是一个递减失效率(DFR)随机变量，如果

$$\lambda(t)\downarrow t\quad（关于\ t\ 递减）$$

如果我们把 X 当作某种元件的寿命，那么因为 $\lambda(t)dt$ 表示历时 t 的元件在区间 $(t,t+dt)$ 内失效的概率，所以可看出，X 是 IFR(DFR)意味着在一小段时间 dt 内，越是老的元件越容易(越不容易)失效。

现在设这种元件到时刻 t 没有坏，以 X_t 记从 t 开始元件的剩余寿命。X_t 的分布 $\overline{F}_t$ 为

(8.1.2)
$$\begin{aligned}\overline{F}_t(a)&=P\{X_t>a\}\\&=P\{X-t>a\mid X>t\}\\&=\overline{F}(t+a)/\overline{F}(t)\end{aligned}$$

命题 8.1.3

$$X\text{ 是 IFR}\Leftrightarrow X_t\text{ 关于 } t \text{ 随机递减}$$
$$X\text{ 是 DFR}\Leftrightarrow X_t\text{ 关于 } t \text{ 随机递增}$$

证明 能够证明 X_t 的失效率函数，记作 λ_t，为

(8.1.3)
$$\lambda_t(a)=\lambda(t+a)$$

其中 λ 是 X 的失效率函数。利用(8.1.2)能正式地证明等式(8.1.3)，或者能较直观地作如下论证：

$$\begin{aligned}\lambda_t(a)&=\lim_{h\to0}P\{a<X_t<a+h\mid X_t\geqslant a\}/h\\&=\lim_{h\to0}P\{a<X-t<a+h\mid X\geqslant t,X-t\geqslant a\}/h\end{aligned}$$

$$= \lim_{h \to 0} P\{t + a < X < t + a + h | X \geqslant t + a\}/h$$

$$= \lambda(t + a)$$

因为

(8.1.4)
$$\overline{F}_t(s) = \exp\left\{-\int_0^s \lambda_t(a)da\right\} = \exp\left\{-\int_t^{t+s} \lambda(y)dy\right\}$$

所以，如果 $\lambda(y)$ 是递增(递减)的，那么 $\overline{F}_t(s)$ 关于 t 是递减(递增)的。类似地，如果 $\overline{F}_t(s)$ 关于 t 是递减(递增)的，那么(8.1.4)蕴含了 $\lambda(y)$ 关于 y 是递增(递减)的。

于是一种元件的寿命是IFR(DFR)，如果这种元件越老，其剩余寿命越是随机地小(大)。

一类常见的DFR分布是指数分布的混合。我们说分布 F 是分布 $F_\alpha(0<\alpha<\infty)$ 的一个混合，如果对某分布 G

$$F(x) = \int_0^\infty F_\alpha(x)dG(\alpha)$$

当我们从一个由不同的类型组成的总体中取样时，就出现混合。一个特征为 α 的类型的元件的值服从分布 F_α。G 是特征量的分布。

现在考虑两个具有失效率 λ_1 和 λ_2 的指数分布的混合，其中 $\lambda_1 < \lambda_2$。为证明这混合分布是DFR，注意到如果选出的元件到时刻 t 没有坏，那么它的剩余寿命的分布仍是两个指数分布的混合。这是因为当它是1-型元件时，它的剩余寿命仍服从失效率为 λ_1 的指数分布，或者当它是2-型元件时，它的剩余寿命仍服从失效率为 λ_2 的指数分布。不过，元件为1-型的概率不再是(先验)概率 p，而是在到时刻 t 没坏的条件下的条件概率。事实上，它是一个1-型元件的概率为

$$P\{1\text{-型} \mid \text{寿命} > t\} = \frac{P\{1\text{-型}, \text{寿命} > t\}}{P\{\text{寿命} > t\}} = \frac{pe^{-\lambda_1 t}}{pe^{-\lambda_1 t} + (1-p)e^{-\lambda_2 t}}。$$

因为上式关于 t 是递增的，所以 t 越大，所用的元件越可能是1-型

的（较好的一种，因为 $\lambda_1<\lambda_2$）。因此，越是老的元件越少可能失效，于是指数分布的混合是 DFR。

DFR 分布函数类在混合下原来是封闭的（这蕴含有上面的事实，因为指数分布有常数失效率，所以既是 IFR 又是 DFR）。为证明这事实，我们需要下列著名的引理，其证明留作练习。

引理 8.1.4 柯西—许瓦兹不等式

对任意分布 G 和函数 $h(t)$，$k(t)$，$t\geqslant 0$，有

$$\left(\int h(t)k(t)dG(t)\right)^2\leqslant\left(\int h^2(t)dG(t)\right)\left(\int k^2(t)dG(t)\right)$$

假定这些积分存在。

现在我们可以叙述下列

命题 8.1.5

如果对一切 $0<\alpha<\infty$，F_α 是 DFR 分布，G 是一个 $(0,\infty)$ 上的分布函数，那么 F 是 DFR，其中

$$F(t)=\int_0^\infty F_\alpha(t)dG(\alpha)$$

证明

$$\lambda_F(t)=\frac{\frac{d}{dt}F(t)}{\overline{F}(t)}=\frac{\int_0^\infty f_\alpha(t)dG(\alpha)}{\overline{F}(t)}$$

我们先设所有的导数存在，然后证明 $(d/dt)\lambda_F(t)\leqslant 0$，这就证明 $\lambda_F(t)$ 关于 t 是递减的。今有

$$\frac{d}{dt}[\lambda_F(t)]=\frac{\overline{F}(t)\int f'_\alpha(t)dG(\alpha)+\left(\int f_\alpha(t)dG(\alpha)\right)^2}{\overline{F}^2(t)}$$

因为 $\overline{F}(t)=\int\overline{F}_\alpha(t)dG(\alpha)$，所以由上式可见，为证 $(d/dt)\lambda_F(t)\leqslant 0$，只需证明

(8.1.5)

$$\left(\int f_\alpha(t)dG(\alpha)\right)^2\leqslant\left(\int\overline{F}_\alpha(t)dG(\alpha)\right)\left(\int -f'_\alpha(t)dG(\alpha)\right)$$

令 $h(\alpha)=(\overline{F}_\alpha(t))^{1/2}$，$k(\alpha)=(-f'_\alpha(t))^{1/2}$，再利用柯西—许瓦兹不等式，我们得到

$$\left(\int(-\overline{F}_\alpha(t)f'_\alpha(t))^{1/2}dG(\alpha)\right)^2\leqslant\int\overline{F}_\alpha(t)dG(\alpha)\int -f'_\alpha(t)dG(\alpha)$$

因此，为了证明(8.1.5)只需证明

(8.1.6)

$$\left(\int f_\alpha(t)dG(\alpha)\right)^2 \leqslant \left(\int(-\overline{F}_\alpha(t)f'_\alpha(t))^{1/2}dG(\alpha)\right)^2$$

由假设 F_α 是 DFR，于是

$$0 \geqslant \frac{d}{dt}\frac{f_\alpha(t)}{\overline{F}_\alpha(t)} = \frac{\overline{F}_\alpha(t)f'_\alpha(t) + f_\alpha^2(t)}{\overline{F}_\alpha^2(t)}$$

这蕴含了

$$-\overline{F}_\alpha(t)f'_\alpha(t) \geqslant f_\alpha^2(t)$$

这就证明了(8.1.6)，由此结论成立。(上式也证明了 $f'_\alpha(t) \leqslant 0$，因此可以定义 $k(\alpha) \equiv (-f'_\alpha(t))^{1/2}$。)

8.2 耦 合

如果 $X \underset{st}{\geqslant} Y$，那么存在与 X 和 Y 同分布的随机变量 X^* 和 Y^*，使得 X^* 以概率 1 大于等于 Y^*。在证明这个结论前我们需要下列引理。

引理 8.2.1

设 F 和 G 是连续的分布函数，如果 X 服从分布 F，那么随机变量 $G^{-1}(F(X))$ 服从分布 G。

证明

$$\begin{aligned} P\{G^{-1}(F(X)) \leqslant a\} &= P\{F(X) \leqslant G(a)\} \\ &= P\{X \leqslant F^{-1}(G(a))\} \\ &= F(F^{-1}(G(a))) \\ &= G(a) \end{aligned}$$

命题 8.2.2

如果分布函数 F 和 G 满足 $\overline{F}(a) \geqslant \overline{G}(a)$，那么存在随机变量 X 和 Y，分别有分布 F 和 G，且使

$$P\{X \geqslant Y\} = 1$$

证明 我们给出 F 和 G 是连续的分布函数时的证明。设 X 服从分布

F,定义 $Y=G^{-1}(F(X))$。那么由引理 8.2.1 知,Y 服从分布 G。但是由于 $F\leqslant G$,故有 $F^{-1}\geqslant G^{-1}$,所以

$$Y = G^{-1}(F(X)) \leqslant F^{-1}(F(X) = X$$

结论得证。

证明 $\overline{F}\geqslant\overline{G}$ 的最简便的方法常常是设 X 是一个具有分布 F 的随机变量,然后用 X 去定义一个随机变量 Y,使得(i)Y 服从分布 G,(ii)$Y\leqslant X$。我们用一些例子来说明这种方法,它以耦合著称。

例 8.2(a) 向量的随机序。 设 $X_1,\cdots,X_n$ 相互独立,$Y_1,\cdots,Y_n$ 相互独立。如果 $X_i\underset{\text{st}}{\geqslant}Y_i$,那么对任意递增函数 f 有

$$f(X_1,\cdots,X_n)\underset{\text{st}}{\geqslant}f(Y_1,\cdots,Y_n)$$

证明 设 $X_1,\cdots,X_n$ 相互独立,然后利用命题 8.2.2 产生相互独立的 $Y_1^*,\cdots,Y_n^*$,其中 Y_i^* 服从 Y_i 的分布,且 $Y_i^*\leqslant X_i$。由于 f 是递增的,所以 $f(X_1,\cdots,X_n)\geqslant f(Y_1^*,\cdots,Y_n^*)$。因此对任意的 a 有

$$f(Y_1^*,\cdots,Y_n^*) > a \Rightarrow f(X_1,\cdots,X_n) > a$$

所以

$$P\{f(Y_1^*,\cdots Y_n^*) > a\} \leqslant P\{f(X_1,\cdots,X_n) > a\}$$

因为上式左边等于 $P\{f(Y_1,\cdots,Y_n)>a\}$,所以结论成立。

例 8.2(b) 泊松随机变量的随机序。 我们证明泊松随机变量关于它的均值随机递增。以 N 记一个均值为 λ 的泊松随机变量。对任意 p,$0<p<1$,令 $I_1,I_2,\cdots$ 相互独立,与 N 独立,且使得

$$I_j = \begin{cases}1, & \text{以概率 } p\\ 0, & \text{以概率 } 1-p\end{cases}$$

那么

$$\sum_{j=1}^{N} I_j \text{ 是均值为 } \lambda p \text{ 的泊松随机变量。(为什么?)}$$

因为

$$\sum_{j=1}^{N} I_j \leqslant N$$

故结论得证。

我们将利用命题 8.1.2 先对向量,然后对随机过程给出随机大于的定义。

定义 我们说随机向量$\underline{X}=(X_1,\cdots,X_n)$随机大于随机向量$\underline{Y}=(Y_1,\cdots,Y_n)$，记为$\underline{X}\underset{\text{st}}{\geqslant}\underline{Y}$，如果对一切递增函数$f$有

$$E[f(\underline{X})]\geqslant E[f(\underline{Y})]$$

我们说随机过程$\{X(t),t\geqslant 0\}$随机大于$\{Y(t),t\geqslant 0\}$，如果对一切$n,t_1,\cdots,t_n$

$$(X(t_1),\cdots,X(t_n))\underset{\text{st}}{\geqslant}(Y(t_1),\cdots,Y(t_n))$$

由例8.2(a)得出，如果$\underline{X}$和$\underline{Y}$是独立分量组成的向量，且满足$X_i\underset{\text{st}}{\geqslant}Y_i$，那么$\underline{X}\underset{\text{st}}{\geqslant}\underline{Y}$。作为练习，请读者给出一个去掉独立性假设时的反例。

在证明一个随机过程随机大于另一个时，耦合又一次经常成为关键。

例 8.2(c) **比较更新过程**。设$N_i=\{N_i(t),t\geqslant 0\},i=1,2$，为更新过程，分别具有来到间隔分布$F$和$G$。如果$\overline{F}\geqslant\overline{G}$，则

$$\{N_1(t),t\geqslant 0\}\underset{\text{st}}{\leqslant}\{N_2(t),t\geqslant 0\}$$

为证明以上结论，我们利用耦合如下。设$X_1,X_2,\cdots$，独立，且分布为F。那么由X_i产生的更新过程(记为N_1^*)与N_1有相同的概率分布。现在产生独立随机变量$Y_1,Y_2,\cdots$，其分布为G，且满足$Y_i\leqslant X_i$。那么由来到间隔时间Y_i产生的更新过程(记为N_2^*)与N_2有相同的概率分布。然而，因为对一切i，有$Y_i\leqslant X_i$，所以推得对一切t

$$N_1^*(t)\leqslant N_2^*(t)$$

这就证明了结论。

8.2.1 生灭过程的随机单调性

设$\{X(t),t\geqslant 0\}$是一个生灭过程。我们将证明$\{X(t),t\geqslant 0\}$的两个随机单调性。第一个是生灭过程随初始状态$X(0)$增加而随机递增。

命题 8.2.3

$\{X(t),t\geqslant 0\}$关于$X(0)$随机递增，即对一切$t_1,\cdots,t_n$和递增函数f，$E[f(X(t_1),\cdots,X(t_n))\mid X(0)=i]$关于$i$递增。

证明 设$\{X_1(t),t\geqslant 0\}$和$\{X_2(t),t\geqslant 0\}$是独立的生灭过程，具有相同

的出生率与死亡率，且设 $X_1(0)=i+1, X_2(0)=i$。因为 $X_1(0)>X_2(0)$，这两个过程始终或增长1或下降1，且它们决不在同一时刻发生增减（因为这种可能性的概率为零），所以要么过程 $X_1(t)$ 一直大于过程 $X_2(t)$，要么在某一时刻二者相等。以 T 记它们首次相等的时刻，即

$$T=\begin{cases}\infty, & \text{若对一切 } t, X_1(t)>X_2(t)\\ inf\{t:X_1(t)=X_2(t)\}, & \text{其它}\end{cases}$$

若 $T<\infty$，则这两个过程在时刻 T 相等。因此，由马尔可夫性知，它们在时刻 T 以后的延续部分具有相同的概率结构。于是，如果我们定义第三个随机过程，记作 $\{X_3(t)\}$，为

$$X_3(t)=\begin{cases}X_1(t), & \text{当 } t<T\\ X_2(t), & \text{当 } t\geqslant T\end{cases}$$

那么 $\{X_3(t)\}$ 也是一个生灭过程，与其它两个过程有相同的参数，而且 $X_3(0)=X_1(0)=i+1$。然而，因为由 T 的定义

$$X_1(t)>X_2(t), \quad \text{对 } t<T$$

可见对一切 t

$$X_3(t)\geqslant X_2(t)$$

这就证明了结论。

第二个随机单调性质说，若初始状态是0，那么在时刻 t 的状态关于 t 随机递增。

命题 8.2.4

对一切 j，$P\{X(t)\geqslant j \mid X(0)=0\}$ 关于 t 递增。

证明 对 $s<t$，

$$\begin{aligned}
&P\{X(t)\geqslant j \mid X(0)=0\}\\
&=\sum_{i=0}^{\infty}P\{X(t)\geqslant j \mid X(0)=0, X(t-s)=i\}P\{X(t-s)=i \mid X(0)=0\}\\
&=\sum_{i=0}^{\infty}P\{X(t)\geqslant j \mid X(t-s)=i\}P_{0i}(t-s) \quad \text{（由可尔可夫性）}\\
&=\sum_{i=0}^{\infty}P\{X(s)\geqslant j \mid X(0)=i\}P_{0i}(t-s)\\
&\geqslant\sum_{i=0}^{\infty}P\{X(s)\geqslant j \mid X(0)=0\}P_{0i}(t-s) \quad \text{（由命题8.2.3）}\\
&=P\{X(s)\geqslant j \mid X(0)=0\}\sum_{i=0}^{\infty}P_{0i}(t-s)
\end{aligned}$$

$$=P\{X(s)\geqslant j|X(0)=0\}$$

注记　命题8.2.4除了提供了生灭过程的转移概率的一个漂亮的定性性质外，在应用中也是有用的。尽管对固定的t，要用显式确定$P_{0j}(t)$的值常常是困难的，但是获得极限概率P_j是件简单的事情。现在由命题8.2.4我们有

$$P\{X(t)\geqslant j|X(0)=0\}\leqslant\lim_{t\to\infty}P\{X(t)\geqslant j|X(0)=0\}=\sum_{i=j}^{\infty}P_i$$

这说明$X(t)$随机小于服从极限分布的随机变量，记为$X(\infty)$，从而提供了$X(t)$的分布的界。

8.2.2　马尔可夫链的指数收敛性

考虑一个有限状态不可约马尔可夫链，转移概率为P_{ij}^n。我们将用耦合方法证明当$n\to\infty$时，P_{ij}^n以指数速度收敛到一个与i无关的极限。为证这一点，我们要利用下述结论：如果一个遍历马尔可夫链具有有限个状态——譬如说M个状态，则必定存在$N,\varepsilon>0$，使得对一切i，j

$$P_{ij}^N>\varepsilon \tag{8.2.1}$$

现在考虑此马尔可夫链的两个独立的拷贝，譬如是$\{X_n, n\geqslant 0\}$和$\{X_n', n\geqslant 0\}$，其中一个从i出发，譬如$P\{X_0=i\}=1$，另一个满足$P\{X_0'=j\}=\pi_j$，$j=1, 2, \cdots, M$，其中π_j是平稳分布，即它们是方程

$$\pi_j=\sum_{i=1}^{M}\pi_i P_{ij},\qquad j=1,\cdots,M$$

$$\sum_{j=1}^{M}\pi_j=1$$

的非负解。

以T记两个过程首次处于同一状态的时刻，即

$$T=\min\{n: X_n=X_n'\}$$

现在

$$T>mN\Rightarrow X_N\neq X_N',\quad X_{2N}\neq X_{2N}',\cdots,X_{mN}\neq X_{mN}'$$

所以

(8.2.2)

$$P\{T>mN\}\leqslant P(A_1)P(A_2|A_1)\cdots P(A_m|A_1\cdots A_{m-1})$$

其中 A_i 为事件 $X_{Ni}\neq X'_{Ni}$。由（8.2.1）得，不管现在的状态如何，在过了时间 N 的将来时刻处于状态 j 的概率至少为 ε，因此不管过去如何，这两个链在过了时间 N 的将来时刻都处于状态 j 的概率至少为 ε^2，从而处于同一状态的概率至少是 $M\varepsilon^2$。因此（8.2.2）右边所有的条件概率不大于$1-M\varepsilon^2$。于是有

(8.2.3) $$P\{T>mN\}\leqslant(1-M\varepsilon^2)^m=(1-\alpha)^m$$

其中 $\alpha\equiv M\varepsilon^2$。

现在我们定义第三个马尔可夫链，记为 $\{\overline{X}_n, n\geqslant 0\}$，它等于 X' 直到 T 为止，之后等于 X，即

$$\overline{X}_n=\begin{cases}X'_n, & \text{对 } n\leqslant T\\ X_n, & \text{对 } n\geqslant T\end{cases}$$

因为 $X'_n=X_T$，所以显然 $\{\overline{X}_n, n\geqslant 0\}$ 是一个转移概率为 P_{ij}的马尔可夫链，且依照平稳分布选取初始状态。今有

$$\begin{aligned}P\{\overline{X}_n=j\}&=P\{\overline{X}_n=j|T\leqslant n\}P\{T\leqslant n\}+P\{\overline{X}_n=j|T>n\}P\{T>n\}\\&=P\{X_n=j|T\leqslant n\}P\{T\leqslant n\}+P\{\overline{X}_n=j,T>n\}\end{aligned}$$

类似地有

$$\begin{aligned}P_{ij}^n&=P\{X_n=j\}\\&=P\{X_n=j|T\leqslant n\}P\{T\leqslant n\}+P\{X_n=j,T>n\}\end{aligned}$$

从而

$$P_{ij}^n-P\{\overline{X}_n=j\}=P\{X_n=j,T>n\}-P\{\overline{X}_n=j,T>n\}$$

这蕴含了

$$\begin{aligned}|P_{ij}^n-P\{\overline{X}_n=j\}|&\leqslant P\{T>n\}\\&\leqslant(1-\alpha)^{n/N-1}\quad(\text{由}(8.2.3))\end{aligned}$$

但容易验证(譬如对 n 用归纳法)

$$P\{\overline{X}_n=j\}=\pi_j$$

于是可见

$$|P_{ij}^n - \pi_j| \leqslant \frac{\beta^n}{1-\alpha}, \qquad \text{其中 } \beta = (1-\alpha)^{1/N}$$

因此 P_{ij}^n 确实以指数速度收敛到与 i 无关的极限。(以上也附带证明了不存在多于一个的平稳分布。)

注记　在8.3节中，我们将用类似于上述定理中给出的论证方法去证明更新过程来到间隔分布连续时的布莱克威尔定理。

8.3　失效率序及其对计数过程的应用

随机变量 X 有一个比 Y 大的失效率函数，如果对一切 $t \geqslant 0$

(8.3.1)　　$$\lambda_X(t) \geqslant \lambda_Y(t)$$

这里 $\lambda_X(t)$ 和 $\lambda_Y(t)$ 分别是 X 和 Y 的失效率函数。(8.3.1)意为，寿命为 X 的元件比寿命为 Y 的元件在年龄相同时前者更可能立即失效。事实上，因为

$$P\{X > t+s | X > t\} = \exp\{-\int_t^{t+s} \lambda(y)dy\}$$

所以 (8.3.1) 等价于

$$P\{X > t+s | X > t\} \leqslant P\{Y > t+s | Y > t\}$$

或等价地，对一切 $t \geqslant 0$

$$X_t \underset{\text{st}}{\leqslant} Y_t$$

其中 X_t 和 Y_t 分别是与 X 和 Y 同分布的元件经历 t 时间后的剩余寿命。

失效率序在比较计数过程时非常有用。为了说明这一点，让我们从考虑一个延迟更新过程开始，这个更新过程的首次更新服从分布 G，而其它的来到间隔服从分布 F，F 和 G 都是连续的，且有失效率函数 $\lambda_F(t)$ 和 $\lambda_G(t)$。令 $\mu(t)$ 满足

$$\max(\max_{0 \leqslant s \leqslant t} \lambda_F(s), \max_{0 \leqslant s \leqslant t} \lambda_G(s)) \leqslant \mu(t)$$

我们先表明如何由取自强度函数为 $\mu(t)$ 的非齐次泊松过程的一

个随机样本产生这个延迟更新过程。

以 $S_1, S_2, \cdots$ 记强度函数为 $\mu(t)$ 的非齐次泊松过程 $\{N(t), t \geqslant 0\}$ 中事件发生的时刻。现在我们来定义一个计数过程，其事件只可能在时刻 S_1，S_2，$\cdots$ 发生，然后论证这个计数过程是一个具有初始更新分布 G 和来到间隔分布 F 的延迟更新过程。令

$$I_i = \begin{cases} 1, & \text{若这个计数过程的一个事件在时刻 } S_i \text{ 发生} \\ 0, & \text{其它} \end{cases}$$

因此，为了确定这个计数过程，我们需要规定 I_i，$i \geqslant 1$，的联合分布。作法如下：

给定 S_1，S_2，$\cdots$，取

(8.3.2) $$P\{I_1 = 1\} = \lambda_G(S_1)/\mu(S_1)$$

且对 $i > 1$，

(8.3.3) $$P\{I_i = 1 | I_1, I_2, \cdots, I_{i-1}\} = \begin{cases} \dfrac{\lambda_G(S_i)}{\mu(S_i)}, & \text{若 } I_1 = \cdots = I_{i-1} = 0 \\ \dfrac{\lambda_F(S_i - S_j)}{\mu(S_i)}, & \text{若 } j = \max\{k : k < i, I_k = 1\} \end{cases}$$

为了对上述定义有点感性认识，以 $A(t)$ 记这个计数过程在时刻 t 的年龄；也就是从这个计数过程在 t 以前发生的最后一个事件到 t 的这段时间。那么 $A(S_1) = S_1$，而其它的 $A(S_i)$ 的值由 I_1，$\cdots$，I_{i-1} 递推地得到。例如，如果 $I_1 = 0$，则 $A(S_2) = S_2$，而当 $I_1 = 1$，则 $A(S_2) = S_2 - S_1$。那么（8.3.2）和（8.3.3）等价于

$$P\{I_i = 1 | I_1, \cdots, I_{i-1}\} = \begin{cases} \dfrac{\lambda_G(S_i)}{\mu(S_i)}, & \text{若 } A(S_i) = S_i \\ \dfrac{\lambda_F(A(S_i))}{\mu(S_i)}, & \text{若 } A(S_i) < S_i \end{cases}$$

我们说由 $I_i, i \geqslant 1$，定义的这个计数过程构成了所期待的延迟更新过程。为了明瞭这一点，注意到对于这个计数过程在已知任一时刻 t 过去的情况的条件下，在时刻 t 发生一个事件的概率强度为

$P\{$在$(t,t+h)$中发生事件$|$到t为止的历史$\}$

$=P\{$非齐次泊松过程的一个事件在$(t,t+h)$中发生，

且被计数$|$到t为止的历史$\}$

$=(\mu(t)h+o(h))P\{$事件被计数$|$到t为止的历史$\}$

$$=\begin{cases}[\mu(t)h+o(h)]\dfrac{\lambda_G(t)}{\mu(t)}=\lambda_G(t)h+o(h), & \text{若 } A(t)=t \\ [\mu(t)h+o(h)]\dfrac{\lambda_F(A(t)}{\mu(t)}=\lambda_F(A(t)h+o(h), & \text{若 } A(t)<t\end{cases}$$

因此在任一时刻t发生一个事件的概率（强度）只依赖于那个时刻的年龄，且年龄为t时，它等于$\lambda_G(t)$，否则等于$\lambda_F(A(t))$。而这样的计数过程显然是一个延迟更新过程，它具有来到间隔分布F和初始更新分布G。

现在我们利用这个延迟更新过程可表为一个非齐次泊松过程的随机取样的事实，给出来到间隔分布连续时的布莱克威尔定理的一个简单的概率证明。

定理（布莱克威尔定理）。设$\{N^*(t),\ t\geqslant 0\}$为一个具有连续来到间隔分布$F$的更新过程，则$t\to\infty$时

$$m(t+a)-m(t)\to\frac{a}{\mu}$$

其中$m(t)=E[N^*(t)]$，μ是来到间隔的平均时间，假设为有限。

证明 我们将在附加的简化假设下证明布莱克威尔定理，假设F的失效率函数$\lambda_F(t)$有界且大于一个正数，即存在$0<\lambda_1<\lambda_2<\infty$，使得对一切$t$

$$\lambda_1<\lambda(t)<\lambda_2 \tag{8.3.4}$$

又令G是一个分布，它的失效率函数也在λ_1和λ_2之间。

考虑一个参数为λ_2的泊松过程，它的事件发生的时刻记为S_1，S_2，…。现在设I_1^*，I_2^*，…由（8.3.3）产生，其中的$\mu(t)\equiv\lambda_2$。设I_1，I_2，…，由（8.3.3）产生，其中的$G=F$和$\mu(t))\equiv\lambda_2$，且与序列I_1^*、I_2^*，…条件独立（给定S_1，S_2，…）。于是，事件在使$I_i^*=1$的时刻S_i发生的计数过程，记为$\{N_0(t),\ t\geqslant 0\}$，是一个来到间隔分布为$F$的延迟更新过程；而事件在使$I_i=1$的时刻$S_i$发生的计数过程，记为$\{N(t),\ t\geqslant 0\}$，是一个来到间隔分布为$F$的更新过程。令

$$N=\min\{i:\ I_i=I_i^*=1\}$$

即 N 是第一个同时被两个新产生的过程计数的泊松过程的事件。因为每个泊松事件与任何其它事情独立地，至少以概率 λ_1/λ_2 被一个指定的新产生的过程计数，所以得

$$P\{I_i=I_i^*=1\mid I_1,\cdots,I_{i-1},I_i^*,\cdots,I_{i-1}^*\}\geqslant\left(\frac{\lambda_1}{\lambda_2}\right)^2$$

从而

$$P\{N<\infty\}=1$$

现在定义第三个序列 $\bar{I}_i$，$i\geqslant 1$，为

$$\bar{I}_i=\begin{cases}I_i^*, & \text{对 } i\leqslant N\\ I_i, & \text{对 } i\geqslant N\end{cases}$$

于是，事件在使 $\bar{I}_i=1$的时刻 S_i 发生的计数过程 $\{\bar{N}(t),\ t\geqslant 0\}$ 是一个具有初始更新分布 G 和来到间隔分布 F 的延迟更新过程，且从时刻 S_N 开始，它的事件发生的时间与$\{N(t),t\geqslant 0\}$的相同。

令 $N(t,t+a)=N(t+a)-N(t)$，对 $\bar{N}$ 也类似地定义。我们有

$$\begin{aligned}E[\bar{N}(t,t+a)]&=E[\bar{N}(t,t+a)\mid S_N\leqslant t]P\{S_N\leqslant t\}\\&\quad+E[\bar{N}(t,t+a)\mid S_N>t]P\{S_N>t\}\\&=E[N(t,t+a)\mid S_N\leqslant t]P\{S_N\leqslant t\}\\&\quad+E[\bar{N}(t,t+a)\mid S_N>t]P\{S_N>t\}\\&=E[N(t,t+a)]+(E[\bar{N}(t,t+a)\mid S_N>t]\\&\quad-E[N(t,t+a)\mid S_N>t]P\{S_N>t\}\end{aligned}$$

现在容易推出 $E[N(t,t+a)\mid S_N>t]\leqslant\lambda_2 a$，以及用 $\bar{N}$ 代替 N 也有类似的估计式。因为 $N<\infty$ 蕴含，当 $t\to\infty$ 时，$P\{S_N>t\}\to 0$，故从上述讨论可见 $t\to\infty$ 时

(8.3.5) $$E[\bar{N}(t,t+a)]-E[N(t,t+a)]\to 0$$

如果我们取$G=F_e$，其中 $F_e(t)=\int_0^t\bar{F}(y)dy/\mu$，容易证实（见第三章命题3.5.2(i)的证明）$G=F_e$ 时，$E[\bar{N}(t)]=t/\mu$，因此，由(8.3.5)得 $t\to\infty$时

$$E[N(t,t+a)]\to a/\mu$$

这就完成了证明。

我们将发现，由一个泊松过程的随机样本产生一个更新过程的方法在得到来到间隔分布具有递减失效率的更新过程的某些单

调性结果时也是有用的。作为准备，我们再定义一个记号。

定义 对一计数过程$\{N(t),t\geqslant 0\}$和任意的时间集合T，定义$N(T)$为在T中发生的事件数。

我们从一个引理开始。

引理8.3.1

设$N=\{N(t),t\geqslant 0\}$为更新过程，它的来到间隔分布具有递减失效率。又设$N_y=\{N_y(t),t\geqslant 0\}$为延迟更新过程，它的第一个来到间隔有分布$H_y$，其余的来到间隔分布为$F$，其中$H_y$是更新过程$N$在时刻$y$的剩余寿命的分布。(这意味着，可以认为$N_y$是一个与$N$有相同来到间隔分布的更新过程从$y$开始的继续。)这时对任意的时间集合$T_1,\cdots,T_n$，有

$$(N(T_1),\cdots,N(T_n))\underset{\text{st}}{\geqslant}(N_y(T_1),\cdots,N_y(T_n))$$

证明 以$N^*=\{N^*(t),t\leqslant y\}$记一个与$N$独立，但来到间隔分布也为$F$的更新过程在$y$之前的那一段。我们把$N_y$看成是$N^*$在时刻$y$以后的继续。令$A^*(y)$为更新过程$N^*$在时刻$y$的年龄。

考虑一个参数为$\mu=\lambda(0)$的泊松过程，以$S_1,S_2,\cdots$记其事件发生的时刻。利用这个泊松过程产生一个计数过程，记为N，在这个过程中，事件只可能在时刻$S_i(i\geqslant 1)$发生。如果当一个事件在S_i发生时，有$I_i=1$，否则$I_i=0$，那么令

$$P\{I_i=1\mid I_1,\cdots\cdots,I_{i-1}\}=\lambda(A(S_i))/\mu$$

其中$A(S_1)=S_1$，而对$i>1$，$A(S_i)$是从S_i之前最后一个被计数的事件到S_i的时间，这里在S_i发生的泊松事件称为被计数的，如果$I_j=1$。如前，这样产生的计数过程是一个来到间隔分布为F的更新过程。

现在我们定义另一个计数过程，它也只可能在时刻$S_i,i\geqslant 1$，发生事件，而以$\bar{I}_i$指明在S_i事件是否发生。以$\bar{A}(t)$记这个过程从t之前的最后一个事件到t的时间，或若t之前没有事件，则定义$\bar{A}(t)$为$t+A^*(y)$。令$\bar{I}_i$使得

如果$I_i=0$，则$\bar{I}_i=0$，

$$\text{如果 } I_i=1\text{，则 } \bar{I}_i=\begin{cases}1, & \text{以概率 } \lambda(\bar{A}(S_i))/\lambda(A(S_i))\\ 0, & \text{其它}\end{cases}$$

因为仅当$I_i=1$时，$\bar{I}_i$才可能等于1，所以我们始终有$\bar{A}(t)\geqslant A(t)$，又因为λ不

增，所以 $\lambda(\bar{A}(S_i))\leqslant\lambda(A(S_i))$，因此以上定义是有意义的。由此可见

$$\begin{aligned}&P\{\bar{I}_i=1|\bar{I}_1,\cdots,\bar{I}_{i-1},I_1,\cdots,I_{i-1}\}\\&=P\{I_i=1|I_1,\cdots,I_{i-1}\}P\{\bar{I}_i=1|I_i=1\}\\&=\lambda(\bar{A}(S_i))/\mu\end{aligned}$$

且知由 $\bar{I}_i,i\geqslant1$，产生的计数过程，记为 N_y，是一个延迟更新过程(因为 $\bar{A}(0)=A^*(y)$，其初始分布为 H_y。因为 N_y 的事件只能在 N 的事件发生的时刻上发生(对一切 $i,\bar{I}_i\leqslant I_i$)，所以对所有的集 T

$$N(T)\geqslant N_y(T)$$

引理得证。

命题8.3.2　DFR 更新过程的单调性

设 $A(t)$ 和 $\bar{Y}(t)$ 为一个更新过程 $N=\{N(t),t\geqslant0\}$ 在时刻 t 的年龄和剩余寿命，其来到间隔分布是 DFR，则 $A(t)$ 和 $\bar{Y}(t)$ 都关于 t 随机递增。也即对一切 a，$P\{A(t)>a\}$ 和 $P\{Y(t)>a\}$ 都关于 t 递增。

证明　设想我们需要证明

$$P\{A(t+y)>a\}\geqslant P\{A(t)>a\}$$

为此我们把 $A(t)$ 当作引理8.3.1中的更新过程 N 在时刻 t 的年龄，把 $A(t+y)$ 当作更新过程 N_y 在时刻 t 的年龄。然后令 $T=[t-a,t]$，由引理8.3.1有

$$P\{N(T)\geqslant1\}\geqslant P\{N_y(T)\geqslant1\}$$

或等价地

$$P\{A(t)\leqslant a\}\geqslant P\{A(t+y)\leqslant a\}$$

对于剩余寿命的证明除了令 $T=[t,t+a]$，其余都是类似的。

命题8.3.2可用来得到 DFR 更新过程的更新函数和 DFR 随机变量的分布的一些良好的界。

系8.3.3

设 F 是一个 DFR 分布，它的前二阶矩为

$$\mu_1=\int xdF(x),\quad \mu_2=\int x^2dF(x)$$

(i)如果 $m(t)$ 是一个来到间隔分布为 F 的更新过程的更新函数，则

$$\frac{t}{\mu_1}\leqslant m(t)\leqslant\frac{t}{\mu_1}+\frac{\mu_2}{2\mu_1^2}-1$$

(ii) $$\overline{F}(t)\geqslant\exp\{-\frac{t}{\mu_1}-\frac{\mu_2}{2\mu_1^2}+1\}$$

证明 (i)以 $X_1,X_2,\cdots$ 记一个来到间隔分布为 F 的更新过程$\{N(t),t\geqslant 0\}$的来到间隔时间。现在

$$\sum_{i=1}^{N(t)+1}X_i=t+Y(t)$$

其中 $Y(t)$是在时刻 t 的剩余寿命。取期望且利用瓦尔德等式，我们有

$$\mu_1(m(t)+1)=t+E[Y(t)]$$

而由命题8.3.2，$E[Y(t)]$关于 t 递增，又因为 $E[Y(0)]=\mu_1$，且(见第三章命题3.4.6)

$$\lim_{t\to\infty}E[Y(t)]=\frac{\mu_2}{2\mu_1}$$

可见

$$t+\mu_1\leqslant\mu_1(m(t)+1)\leqslant t+\frac{\mu_2}{2\mu_1}$$

或

$$\frac{t}{\mu_1}\leqslant m(t)\leqslant\frac{t}{\mu_1}+\frac{\mu_1}{2\mu_1^2}-1$$

(ii) 对恒等式$m(t)=\sum_{n=1}^{\infty}F_n(t)$求微分得

$$\begin{aligned}m'(t)dt&=\sum_{n=1}^{\infty}F_n'(t)dt\\&=\sum_{n=1}^{\infty}P\{\text{第}n\text{ 次更新发生在}(t,t+dt)\text{之中}\}+o[(dt)]\\&=P\{\text{在}(t,t+dt)\text{中发生一次更新}\}+o[(dt)]\end{aligned}$$

于是 $m'(t)$等于在时刻 t 发生一次更新的概率(强度)。但因为 $\lambda(A(t))$是已知直到 t 为止的历史的条件下，在时刻 t 发生一次更新的概率(强度)，所以我们有

$$\begin{aligned}m'(t)&=E[\lambda(A(t))]\\&\geqslant\lambda(t)\end{aligned}$$

而上述不等式成立是因为 λ 递减，且 $A(t)\leqslant t$。对上述不等式积分得

$$m(t) \geqslant \int_0^t \lambda(s) ds$$

由于

$$\overline{F}(t) = \exp\left(-\int_0^t \lambda(s) ds\right)$$

可见

$$\overline{F}(t) \geqslant e^{-m(t)}$$

然后由(i)得结论。

8.4 似然比序

设 X 和 Y 为连续、非负的随机变量，分别有密度 f 和 g。我们说 X 在似然比意义下大于 Y，记为

$$X \underset{\mathrm{LR}}{\geqslant} Y$$

如果对一切 $x \leqslant y$

$$\frac{f(x)}{g(x)} \leqslant \frac{f(y)}{g(y)}$$

因此 $X \underset{\mathrm{LR}}{\geqslant} Y$，如果它们各自密度的比 $f(x)/g(x)$ 是 x 的非降函数。我们先注意到，似然比序强于失效率序（它又强于随机大于序）。

命题8.4.1

设 X 和 Y 是非负随机变量，有密度 f 和 g，失效率 λ_X 和 λ_Y。如果

$$X \underset{\mathrm{LR}}{\geqslant} Y$$

则对一切 $t \geqslant 0$，有

$$\lambda_X(t) \leqslant \lambda_Y(t)$$

证明 因为 $X \underset{\mathrm{LR}}{\geqslant} Y$，故对 $x \geqslant t$，有 $f(x) \geqslant g(x)f(t)/g(t)$。从而

$$\lambda_X(t) = \frac{f(t)}{\int_t^\infty f(x) dx}$$

$$\leqslant \frac{f(t)}{\int_t^\infty g(x) f(t)/g(t) dx}$$

$$=\frac{g(t)}{\int_t^{\infty} g(x)dx}$$

$$=\lambda_Y(t)$$

例8.4(a) 若 X 是参数为 λ 的指数随机变量，Y 是参数为 μ 的指数随机变量，则

$$\frac{f(x)}{g(x)}=\frac{\lambda}{\mu}e^{(\mu-\lambda)x}$$

故当 $\lambda\leqslant\mu$ 时，有 $X\underset{LR}{\geqslant}Y$。

例8.4(b) **一个统计推断问题**。统计中的一个中心问题，是对一个给定的随机变量的未知分布作出推断。在最简单的情形，我们设 X 为一个连续随机变量，它有密度或为 f 或为 g。根据 X 的观察值，我们必须决定是 f 还是 g。

上述问题的一个决策规划是一个函数 $\phi(x)$，它或取值0或取值1，其意义是若 X 的观察值为 x，$\phi(x)=0$时决定密度是 f，$\phi(x)=1$时则为 g。为了有助于我们作出一个好的决策规则，先注意到

$$\int_{x:\phi(x)=1} f(x)dx=\int f(x)\phi(x)dx$$

表示当 f 是真实的密度时，拒绝 f 的概率。获得一个决策规则的经典方法是固定一个常数 α，$0\leqslant\alpha\leqslant1$，然后限于考虑满足

(8.4.1) $$\int f(x)\phi(x)dx\leqslant\alpha$$

的决策规划。在这些规则中要选出一个，它使得当 f 不是真实的密度时，拒绝 f 的概率最大。也即，它使

$$\int_{x:\phi(x)=1} g(x)dx=\int g(x)\phi(x)dx$$

最大。按此准则的最优决策规则由下面的命题给出，它以奈曼—皮尔逊引理著称。

奈曼—皮尔逊引理

所有满足（8.4.1）的决策规则 ϕ 中，使 $\int g(x)\phi(x)dx$ 最大的 ϕ^* 为

$$\phi^*(x)=\begin{cases}0, & 若\ f(x)/g(x)\geqslant c\\ 1, & 若\ f(x)/g(x)<c\end{cases}$$

其中 c 选取得使

$$\int f(x)\phi^*(x)dx=\alpha$$

证明 设 ϕ 满足(8.4.1)，对任一 x

$$(\phi^*(x)-\phi(x))(cg(x)-f(x))\geqslant 0$$

因为 $\phi^*(x)=1$时，乘积的两项均非负，而 $\phi^*(x)=0$时，两项均非正，所以上式成立。从而有

$$\int(\phi^*(x)-\phi(x))(cg(x)-f(x))dx\geqslant 0$$

故

$$\begin{aligned}
&c\left[\int\phi^*(x)g(x)dx-\int\phi(x)g(x)dx\right]\\
&\geqslant\int\phi^*(x)f(x)dx-\int\phi(x)f(x)dx\\
&\geqslant 0
\end{aligned}$$

这就证明了结论。

如果我们现在假设 f 和 g 有单调似然比序，即$f(x)/g(x)$关于 x 非降，那么这个最优决策规则可写成

$$\phi^*(x)=\begin{cases}0, & \text{当 } x\geqslant k\\ 1, & \text{当 } x<k\end{cases}$$

其中 k 满足

$$\int_{-\infty}^{k}f(x)dx=\alpha$$

也就是说，这个最优决策规则是，当观察值大于某一临界值时，则决定真实密度为 f，否则决定为 g。

似然比序在最优化理论中有着重要的应用。下列命题是十分有用的。

命题8.4.2

设 X 和 Y 相互独立，分别具有密度 f 和 g，且设

$$X\underset{LR}{\geqslant}Y$$

如果 $h(x,y)$是一个实值函数，满足：$x\geqslant y$ 时

$$h(x,y) \geqslant h(y,x)$$

则

$$h(X,Y) \underset{\text{st}}{\geqslant} h(Y,X)$$

证明 令 $U=\max(X,Y)$，$V=\min(X,Y)$。那么在条件 $U=u$，$V=v$，$u \geqslant v$ 之下，$h(X,Y)$ 的条件分布集中在 $h(u,v)$ 及 $h(v,u)$ 两点上，且取较大的值 $h(u,v)$ 的概率为

$$\begin{aligned}\lambda_1 &\equiv P\{X=\max(X,Y),Y=\min(X,Y)\mid U=u,V=v\}\\ &=\frac{f(u)g(v)}{f(u)g(v)+f(v)g(u)}\end{aligned}$$

类似地有，在条件 $U=u$，$V=v$ 之下，$h(Y,X)$ 也集中在两点 $h(u,v)$ 及 $h(v,u)$ 上，且取值 $h(u,v)$ 的概率为

$$\begin{aligned}\lambda_2 &\equiv P\{Y=\max(X,Y),X=\min(X,Y)\mid U=u,V=v\}\\ &=\frac{g(u)f(v)}{g(u)f(v)+f(u)g(v)}\end{aligned}$$

因为 $u \geqslant v$

$$f(u)g(v) \geqslant g(u)f(v)$$

故，在条件 $U=u$ 和 $V=v$ 之下，$h(X,Y)$ 随机大于 $h(Y,X)$，即

$$P\{h(X,Y) \geqslant a \mid U,V\} \geqslant P\{h(Y,X) \geqslant a \mid U,V\}$$

现在对上式两边取期望，就得结论。

注记 或许有点令人惊奇，当我们只假设 $X \underset{st}{\geqslant} Y$ 时，上述结论不一定成立。作为一个反例，注意到 $x \geqslant y$ 时 $2x+y \geqslant x+2y$。然而，若 X 和 Y 是独立的，且

$$X=\begin{cases}3, & \text{以概率}0.2\\ 9, & \text{以概率}0.8\end{cases}$$

$$Y=\begin{cases}1, & \text{以概率}0.2\\ 4, & \text{以概率}0.8\end{cases}$$

则 $X \underset{\text{st}}{\geqslant} Y$，但 $P\{2X+Y \geqslant 11\}=0.8$，而 $P\{2Y+X \geqslant 11\}=0.8+(0.2)(0.8)=0.96$。于是 $2X+Y$ 不随机大于 $2Y+X$。

命题8.4.2在最优排序问题中有重要的应用。例如，设 n 个工件要排成某种加工顺序，每个工件有某种可度量的特征，例如一个工件的可度量的特征可以是完成这工件的加工所需时间。假设，

若 x_i 是第 i 个工件的可度量特征，所选的加工顺序是1，2，…，n 的一个排列 i_1，i_2，…，i_n，那么报酬为 $h(x_{i_1}, x_{i_2}\cdots, x_{i_n})$。现在我们设第 i 个工件的可度量特征是一个随机变量，譬如说是 X_i，$i=1$，…，n。如果

$$X_1 \underset{\text{LR}}{\geqslant} X_2 \underset{\text{LR}}{\geqslant} \cdots \underset{\text{LR}}{\geqslant} Y_n$$

且若 h 满足 $y_i > y_{i-1}$ 时

$$h(y_1, \cdots, y_{i-1}, y_i, y_{i+1}, \cdots, y_n) \geqslant h(y_1, \cdots, y_i, y_{i-1}, y_{i+1}, \cdots, y_n)$$

那么由命题8.4.2得出，顺序1，2，…，n（n，$n-1$，…，1）使报酬随机极大（极小）。为了明瞭这一点，考虑任一不从第1个工件开始的顺序，譬如$(i_1, i_2, 1, i_3, \cdots, i_{n-1})$。对 X_{i_1}，X_{i_3}，…，$X_{i_{n-1}}$ 的值取条件，我们可以用命题8.4.2证明，顺序$(i_1, 1, i_2, i_3, \cdots, i_{n-1})$导致一个随机更大的报酬。继续这样的交换就得结论：1，2，…，n 使报酬随机极大。（类似的论证可证明 n，$n-1$，…，1使报酬随机极小。）

具有密度 f 的连续随机变量 X 称为有递增似然比，如果log$(f(x))$是凹的；如果log$f(x)$是凸的，称为有递减似然比。给出这个术语的原因是，随机变量 $c+X$ 有密度 $f(x-c)$，故

$$\begin{aligned} c_1+X \underset{\text{LR}}{\geqslant} c_2+X, &\quad \text{对一切 } c_1 \geqslant c_2 \\ \Leftrightarrow \quad \frac{f(x-c_1)}{f(x-c_2)} \uparrow x, &\quad \text{对一切 } c_1 \geqslant c_2 \\ \Leftrightarrow \quad \log f(x-c_1) - \log f(x-c_2) \uparrow x, &\quad \text{对一切 } c_1 \geqslant c_2 \\ \Leftrightarrow \quad \log f(x) \text{是凹的} & \end{aligned}$$

因此，X 具有递增（递减）似然比，如果 c 增加时，$c+X$ 按似然比序递增（递减）。

作为第二个解释，回想到记号 X_t 表示一个寿命为 X 的元件达到年龄 t 之后的剩余寿命。既然

$$\begin{aligned} \overline{F}_t(a) &\equiv P\{X_t > a\} \\ &= \overline{F}(t+a)/\overline{F}(t) \end{aligned}$$

故 X_t 的密度为

$$f_t(a)=f(t+a)/\overline{F}(t)$$

因此，

$$X_s \underset{\text{LR}}{\geqslant} X_t, \text{对一切 } s\leqslant t \Leftrightarrow \frac{f(s+a)}{f(t+a)}\uparrow a, \quad \text{对一切 } s\leqslant t$$

$$\Leftrightarrow \log f(x) \text{是凹}$$

所以，X 有递增似然比，如果 s 增加时，X_s 按似然比序递减。类似地，X 有递减似然比，如果 s 增加时，X_s 按似然比序递增。

命题8.4.3

若 X 有递增似然比，则 X 是IFR。类似地，若 X 有递减似然比，则 X 是DFR。

证明

$$X_s \underset{\text{LR}}{\geqslant} X_t \Rightarrow \lambda_{X_s} \leqslant \lambda_{X_t} \qquad \text{(由命题8.4.1)}$$

$$\Rightarrow X_s \underset{\text{st}}{\geqslant} X_t$$

注记

(1) 使$\log f(x)$为凹函数的密度函数 $f(x)$ 称为2阶波里亚频率密度。

(2) 也可对值域相同的离散随机变量定义似然比序。如果 $P\{X=x\}/P\{Y=x\}$ 关于 x 递增，则称 $X \underset{\text{LR}}{\geqslant} Y$。

8.5 随机更多变

我们记得一个函数 h 称为凸的，如果对一切$0<\lambda<1$，x_1，x_2，有

$$h(\lambda x_1+(1-\lambda)x_2)\leqslant \lambda h(x_1)+(1-\lambda)h(x_2)$$

我们说 X 比 Y 更多变，记为 $X \underset{\text{v}}{\geqslant} Y$，如果对一切递增凸函数 h

(8.5.1) $$E[h(X)]\geqslant E[h(Y)]$$

若 X 和 Y 分别有分布 F 和 G，那么(8.5.1)成立时，我们也说 $F \underset{\text{v}}{\geqslant} G$。当(8.5.1)成立时为什么我们称 X 比 Y 更多变的解释将推

迟到证明了下述结果之后再作出。

命题8.5.1

若 X 和 Y 是分别具有分布 F 和 G 的非负随机变量，则 $X \underset{v}{\geqslant} Y$ 当且仅当对一切 $a \geqslant 0$

$$\int_a^{\infty} \overline{F}(x)dx \geqslant \int_a^{\infty} \overline{G}(x)dx \tag{8.5.2}$$

证明 先设 $X \underset{v}{\geqslant} Y$。设 h_a 定义为

$$h_a(x) = (x-a)^+ = \begin{cases} 0, & x \leqslant a \\ x-a, & x > a \end{cases}$$

由于 h_a 是一个凸的增函数，所以我们有

$$E[h_a(X)] \geqslant E[h_a(Y)]$$

但是

$$\begin{aligned} E[h_a(X)] &= \int_0^{\infty} P\{(X-a)^+ > x\} dx \\ &= \int_0^{\infty} P\{X > a+x\} dx \\ &= \int_a^{\infty} \overline{F}(y)dy \end{aligned}$$

类似地有

$$E[h_a(Y)] = \int_a^{\infty} \overline{G}(y)dy$$

从而证明了(8.5.2)。为证另一方向的结论，设(8.5.2)对一切$a \geqslant 0$成立，以 h 记一个凸的增函数，我们假设它是二次可微的。因为 h 为凸的等价于 $h'' \geqslant 0$，所以从(8.5.2)得

$$\int_0^{\infty} h''(a) \int_a^{\infty} \overline{F}(x)dxda \geqslant \int_0^{\infty} h''(a) \int_a^{\infty} \overline{G}(x)dxda \tag{8.5.3}$$

计算上式的左边：

$$\begin{aligned} \int_0^{\infty} h''(a) \int_a^{\infty} \overline{F}(x)dxda &= \int_0^{\infty} \int_0^{x} h''(a)da\overline{F}(x)dx \\ &= \int_0^{\infty} h'(x)\overline{F}(x)dx - h'(0)E[X] \\ &= \int_0^{\infty} h'(x) \int_x^{\infty} dF(y)dx - h'(0)E[X] \\ &= \int_0^{\infty} \int_0^{y} h'(x)dxdF(y) - h'(0)E[X] \end{aligned}$$

$$=\int_0^{\infty} h(y)dF(y)-h(0)-h\ (0)E[X]$$

$$=E[h(X)]-h(0)-h'(0)E[X]$$

由于用 $\bar{F}$ 代替 $\bar{G}$ 时，类似的等式也成立，所以由（8.5.3）可知

(8.5.4) $$E[h(X)]-E[h(Y)]\geq h'(0)(E[X]-E[Y])$$

因为 $h'(0)\geq 0$（h 是递增的），且在(8.5.2)中令 $a=0$ 就有 $E[X]\geq E[Y]$，所以上述不等式的右边是非负的。

系8.5.2

若 X 和 Y 是非负随机变量，使得 $E[X]=E[Y]$，则 $X\geq_{v} Y$ 当且仅当对一切凸函数 h

$$E[h(X)]\geq E[h(Y)]$$

证明 设 h 为凸函数，且设 $X\geq_{v} Y$。不等式(8.5.4)是在 h 为凸的假设下得到的，故由于 $E[X]=E[Y]$，它化为

$$E[h(X)]\geq E[h(Y)]$$

结论得证。

因此，对两个具有相同均值的非负随机变量，如果对一切凸函数 h，有 $E[h(X)]\geq E[h(Y)]$，那么 $X\geq_{v} Y$。正是由于这个原因，我们说 $X\geq_{v} Y$ 意味着 X 比 Y 更多变，也就是，直观地看，X 比 Y 更多变，如果它给边端的值更大的权，而保证这一点的一条途径就是要求：h 为凸函数的，有 $E[h(X)]\geq E[h(Y)]$。（例如，由于 $E[X]=E[Y]$ 和 $h(x)=x^2$ 为凸函数，所以我们有 $\mathrm{Var}(X)\geq \mathrm{Var}(Y)$。）

系8.5.3

若 X 和 Y 是非负随机变量，且 $E[X]=E[Y]$，则 $X\geq_{v} Y$ 蕴含 $-X\geq_{v} -Y$。

证明 以 h 记一个递增的凸函数。我们必须证明

$$E[h(-X)]\geq E[h(-Y)]$$

然而，这由系8.5.2可得，因为函数 $f(x)=h(-x)$ 为凸。

我们的下一个结果是处理变动性序的保持问题。

命题8.5.4

若 X_1, X_2, …, X_n 相互独立，Y_1, …, Y_n 相互独立，且 $X_i \underset{v}{\geqslant} Y_i$, $i=1$, …, n，则对一切递增函数 g

$$g(X_1,\cdots,X_n) \underset{v}{\geqslant} g(Y_1,\cdots,Y_n)$$

证明 从假设$2n$ 个随机变量都相互独立开始。对 n 用归纳法证明。当 $n=1$时，我们必须证明，当 g 和 h 是递增凸函数，且$X_1 \underset{v}{\geqslant} Y_1$时，有

$$E[h(g(X_1))] \geqslant E[h(g(Y_1))]$$

这由 $X_1 \underset{v}{\geqslant} Y_1$的定义及 $h(g(x))$是递增凸函数可得，而$h(g(x))$为递增凸函数是因为

$$\frac{d}{dx}h(g(x))=h'(g(x))g'(x) \geqslant 0$$

$$\frac{d^2}{dx^2}h(g(x))=h''(g(x))(g'(x))^2+h'(g(x))g''(x) \geqslant 0$$

设结论对 $n-1$个分量的向量成立，再令 g 和 h 是递增凸函数，现在

$$\begin{aligned}
&E[h(g(X_1,X_2,\cdots,X_n))|X_1=x] \\
&=E[h(g(x,X_2,\cdots\cdots,X_n))|X_1=x] \\
&=E[h(g(x,X_2,\cdots,X_n))] && (\text{由 } X_1,\cdots,X_n \text{ 的独立性}) \\
&\geqslant E[h(g(x,Y_2,\cdots,Y_n))] && (\text{由归纳法假设}) \\
&=E[h(g(X_1,Y_2,\cdots,Y_n))|X_1=x] && (\text{由 } X_1,Y_2,\cdots,Y_n \text{ 的独立性})
\end{aligned}$$

取期望得

$$E[h(g(X_1,X_2,\cdots,X_n))] \geqslant E[h(g(X_1,Y_2,\cdots,Y_n))]$$

然而，对 Y_2, …, Y_n 取条件且利用 $n=1$时的结果，我们能够证明

$$E[h(g(X_1,Y_2,\cdots,Y_n))] \geqslant E[h(g(Y_1,Y_2,\cdots,Y_n))]$$

这就证明了结论。因为假设$2n$ 个随机变量是独立的不影响 $g(X_1, \cdots, X_n)$ 和 $g(Y_1, \cdots, Y_n)$ 的分布，所以在两组 n 个随机变量各自独立的较弱的假设下，结论仍然为真。

8.6 变动性序的应用

在叙述变动性序的一些应用之前，我们将确定一类随机变量，它们比指数随机变量更多(或更少)变。

定义 非负随机变量 X 称为新的平均比用过的好(NBUE)，

如果对一切 $a \geqslant 0$

$$E[X-a|X>a] \leqslant E[X]$$

它称为新的平均比用过的差（NWUE），如果对一切 $a \geqslant 0$

$$E[X-a|X>a] \geqslant E[X]$$

如果我们把 X 看作某种元件的寿命，那么 X 是 NBUE (NWUE)意味着：任一个用过的元件的平均剩余寿命小于(大于)或等于一个新的元件的平均寿命。如果 X 是 NBUE，F 是 X 的分布，则称 F 为一个 NBUE 分布。对 NWUE 也类似地定义。

命题 8.6.1

若 F 是一个 NBUE 分布，均值为 μ，则

$$F \underset{\text{v}}{\leqslant} \exp(\mu)$$

其中 $\exp(\mu)$ 是均值为 μ 的指数分布。若 F 是 NWUE，那么上述不等式反向。

证明 设 F 是均值为 μ 的 NBUE。由命题8.5.1，我们必须证明，对一切 $c \geqslant 0$

$$\int_c^\infty \overline{F}(x)dx \leqslant \int_c^\infty e^{-x/\mu}dx \tag{8.6.1}$$

如果 X 服从分布 F，则

$$\begin{aligned} E[X-a|X>a] &= \int_0^\infty P\{X-a>x|X>a\}dx \\ &= \int_0^\infty \frac{\overline{F}(a+x)}{\overline{F}(a)}dx \\ &= \int_a^\infty \frac{\overline{F}(y)}{\overline{F}(a)}dy \end{aligned}$$

因此，对均值为 μ 的 NBUE 分布 F，有

$$\int_a^\infty \frac{\overline{F}(y)}{\overline{F}(a)}dy \leqslant \mu$$

或

$$\frac{\overline{F}(a)}{\int_a^\infty \overline{F}(y)dy} \geqslant \frac{1}{\mu}$$

它蕴含

$$\int_0^c \left(\frac{\overline{F}(a)}{\int_a^\infty \overline{F}(y)dy} \right) da \geqslant \frac{c}{\mu}$$

我们可以通过变量交换 $x=\int_a^\infty \overline{F}(y)dy, dx=-\overline{F}(a)da$ 求上式左边的值，得

$$-\int_\mu^{x(c)} \frac{dx}{x} \geqslant \frac{c}{\mu}$$

其中 $x(c)=\int_c^\infty \overline{F}(y)dy$。积分得

$$-\log\left(\int_c^\infty \frac{\overline{F}(y)dy}{\mu} \right) \geqslant \frac{c}{\mu}$$

或

$$\int_c^\infty \overline{F}(y)dy \leqslant \mu e^{-c/\mu}$$

这就证明了 (8.6.1)。F 是 NWUE 时证明类似。

8.6.1 比较 $G/G/1$ 排队系统

$G/G/1$ 系统假设顾客来到间隔时间 X_n ($n\geqslant 1$) 是独立同分布，相继的服务时间 S_n ($n\geqslant 1$) 也是如此。只有一个服务员，服务规则为“先来先服务”。

如果我们以 D_n 记第 n 个顾客在排队中的等待时间，那么容易验证（如果你不能验证它，见第七章7.1节）下列递推公式：

$$D_1=0$$

(8.6.2) $$D_{n+1}=\max\{0,\ D_n+S_n-X_{n+1}\}$$

定理 8.6.2 考虑两个 $G/G/1$ 系统。第 i 个系统($i=1,2$)具有来到间隔时间 $X_n^{(i)}$ 和服务时间 $S_n^{(i)}, n\geqslant 1$。以 $D_n^{(i)}$ 记在系统 $i(i=1,2)$ 中的第 n 个顾客在排队中的等待时间。如果有

(i) $E[X_n^{(1)}]=E[X_n^{(2)}]$

(ii) $X_n^{(1)} \underset{v}{\geqslant} X_n^{(2)}$, $S_n^{(1)} \underset{v}{\geqslant} S_n^{(2)}$

那么对一切 n

$$D_n^{(1)} \underset{v}{\geqslant} D_n^{(2)}$$

证明 用归纳法证明。因为对 $n=1$ 它是显然的，所以假设对 n 成立。今有

$$D_n^{(1)} \underset{v}{\geqslant} D_n^{(2)} \qquad \text{(由归纳法假设)}$$
$$S_n^{(1)} \underset{v}{\geqslant} S_n^{(2)} \qquad \text{(由假定)}$$
$$-X_n^1 \underset{v}{\geqslant} -X_n^2 \qquad \text{(由系 8.5.3)}$$

从而，由命题8.5.4，得

$$D_n^{(1)} + S_n^{(1)} - X_{(n+1)}^{(1)} \underset{v}{\geqslant} D_n^{(2)} + S_n^{(2)} - X_{(n+1)}^{(2)}$$

因为 $h(x)=\max(0,x)$是一个递增凸函数，所以由递推公式(8.6.2)和命题8.5.4推出

$$D_{n+1}^{(1)} \underset{v}{\geqslant} D_{n+1}^{(2)}$$

从而完成了证明。

以 $W_Q=\lim E[D_n]$ 记一个顾客在排队中的平均等待时间。

系 8.6.3(对一个 $E[S]<E[X]$的 $G/G/1$排队系统)

(i) 若来到间隔的分布是均值为$1/\lambda$的 NBUE，则

$$W_Q \leqslant \frac{\lambda E[S^2]}{2(1-\lambda E[S])}$$

若分布为 NWUE，则上述不等式反向。

(ii) 若服务时间的分布是均值为$1/\mu$的 NBUE，则

$$W_Q \leqslant \mu\beta(1-\beta)$$

其中 β 是

$$\beta = \int_0^\infty e^{-\mu t(1-\beta)} dG(t)$$

的解，而 G 是来到间隔的分布。若服务时间分布是 NWUE，则上述不等式反向。

证明 由命题8.6.1，一个 NBUE 分布比一个具有相同均值的指数分布较少变。因此，在(i)中我们可以将系统与 $M/G/1$比较，在(ii)中与 $G/M/1$比较。由于(i)和(ii)中不等式的右边分别是 $M/G/1$和 $G/M/1$系统中顾客在排队中的平均等待时间(见第四章例4.3(a)和例4.3(b))，所以由定理8.6.2得出结论。

8.6.2 对更新过程的一个应用

设$\{N_F(t),t\geqslant 0\}$为一个来到间隔分布为 F 的更新过程。在这一节中，我们的目的是证明下列定理。

定理 8.6.4 若 F 是均值为 μ 的 NBUE，则 $N_F(t) \underset{v}{\leqslant} N(t)$，其中 $N(t)$ 是参数为 $1/\mu$ 的泊松过程。若 F 是 NWUE，则不等式反向。

极为有趣的是，证明一个更一般的结论，反而更容易。(关于这一点以后还要谈到。)

引理 8.6.5

令 F_i，$i \geqslant 1$，是均值为 μ 的 NBUE 分布，且 G 是均值为 μ 的指数分布，那么对每个 $k \geqslant 1$

$$\sum_{i=k}^{\infty}(F_1 * F_2 * \cdots * F_i)(t) \leqslant \sum_{i=k}^{\infty} G_{(i)}(t) \tag{8.6.3}$$

其中 $*$ 表示卷积，$G_{(i)}$ 是 G 自身的 i 重卷积。

证明 对 k 用归纳法证明。为证 $k=1$ 时成立，设 $X_1, X_2, \cdots$ 相互独立，X_i 具有分布 F_i，令

$$N^*(t) = \max\{n: \sum_{i=1}^{n} X_i \leqslant t\}$$

现在由瓦尔德等式得

$$E\left[\sum_{i=1}^{N^*(t)+1} X_i\right] = E[X]E[N^*(t) + 1] \tag{8.6.4}$$

能够证明瓦尔德等式对独立的、非负的、具有相同均值的 X_i 也成立(尽管它们不同分布)。我们也有

$$\sum_{i=1}^{N^*(t)+1} X_i = t + \text{从 } t \text{ 直到 } N^* \text{ 增加为止的时间}$$

而 E[从 t 直到 N^* 增加为止的时间]等于诸 X_i 中的一个的平均剩余寿命，于是由 NBUE，它小于或等于 μ，即

$$E\left[\sum_{i=1}^{N^*(t)+1} X_i\right] \leqslant t + \mu$$

然后由(8.6.4)，得

$$E[N^*(t)] \leqslant t/\mu \tag{8.6.5}$$

然而

$$E[N^*(t)] = \sum_{i=1}^{\infty} P\{N^*(t) \geqslant i\}$$

$$= \sum_{i=1}^{\infty} P\{X_1 + X_2 + \cdots + X_i \leqslant t\}$$

$$= \sum_{i=1}^{\infty} (F_1 * \cdots * F_i)(t)$$

类似地，$k=1$时，因为(8.6.3)的右边是指数更新过程在时刻t的均值(或t/μ)，所以，由(8.6.5)知(8.6.3)在$k=1$时成立。

现在设(8.6.3)对k成立，我们有

$$\sum_{i=k+1}^{\infty} (F_1 * \cdots * F_i)(t)$$

$$= \sum_{i=k+1}^{\infty} \int_0^t (F_1 * \cdots * F_{i-1})(t-x) dF_i(x)$$

$$= \int_0^t \sum_{j=k}^{\infty} (F_1 * \cdots * F_j)(t-x) dF_{j+1}(x)$$

$$\leqslant \int_0^t \sum_{j=k}^{\infty} G_{(j)}(t-x) dF_{j+1}(x) \qquad \text{(由归纳法假设)}$$

$$= \sum_{j=k}^{\infty} (G_{(j)} * F_{j+1})(t)$$

$$= \left(\left(\sum_{j=k}^{\infty} G_{(j-1)} * F_{j+1} \right) * G \right)(t)$$

$$\leqslant \left(\left(\sum_{j=k}^{\infty} G_{(j)} \right) * G \right)(t) \qquad \text{(由归纳法假设)}$$

$$= \sum_{j=k}^{\infty} G_{(j+1)}(t)$$

$$= \sum_{i=k+1}^{\infty} G_{(i)}(t)$$

这就完成了证明。

现在我们已准备好证明定理8.6.4。

定理8.6.4的证明 由命题8.5.1，我们必须证明对一切$k \geqslant 1$

$$\sum_{j=k}^{\infty} P\{N_F(t) \geqslant i\} \leqslant \sum_{j=k}^{\infty} P\{N(t) \geqslant i\}$$

但是，上式的左边等于(8.6.3)的左边，其中每个F_i等于F，而上式的右边正是(8.6.3)的右边。

注记 倘若我们试图直接证明

$$\sum_{j=k}^{\infty} F_{(i)}(t) \leqslant \sum_{j=k}^{\infty} G_{(i)}(t)$$

其中 F 是 NBUE，G 是指数分布，具有相同的均值。那么我们可能试图对 k 用归纳法。$k=1$ 时的证明与我们在引理8.6.5中所用完全一样。可是，当我们从假设结论对 k 成立去证明对 $k+1$ 也成立时，我们会证到这样一步

$$\sum_{i=k+1}^{\infty} F_{(i)} \leqslant \left(\sum_{j=k}^{\infty} G_{(j-1)} * F \right) * F$$

然而，在这一步，归纳法假设还不足以使我们推断出右边小于或等于 $\left(\sum_{j=k}^{\infty} G_{(j)} \right) * F$。道理在于，有时使用归纳法时证明一个更强的结论反而更加容易，因为归纳法假设也提供了更多的东西可派用处。

8.6.3 对分支过程的一个应用*

考虑两个分支过程，以 F_1 和 F_2 分别记这两个过程中的每个个体的后代数的分布。假设

$$F_1 \underset{v}{\geqslant} F_2$$

也就是，我们假设在第一个过程中每个个体的后代数更多变，以 $Z_n^{(i)}(i=1,2)$ 记第 i 个过程的第 n 代的总量。

定理 8.6.6 若 $Z_0^{(i)}=1$，$i=1,2$，则对一切 n，$Z_n^{(1)} \underset{v}{\geqslant} Z_n^{(2)}$。

证明 对 n 用归纳法证明。因为它对 $n=0$ 成立，所以设它对 n 成立。今有

$$Z_{n+1}^{(1)} = \sum_{j=1}^{Z_n^{(1)}} X_j$$

和

$$Z_{n+1}^{(2)} = \sum_{j=1}^{Z_n^{(2)}} Y_j$$

其中 X_j 是独立的，具有分布 F_1（X_j 表示第一个过程的第 n 代中的第 j 个人

---

* 为复习这种模型，读者应参考第四章4.5节。

的后代数)，Y_j 是独立的，具有分布 F_2。因为

$$X_j \underset{v}{\geqslant} Y_j \qquad \text{(由假定)}$$

且

$$Z_n^{(1)} \underset{v}{\geqslant} Z_n^{(2)} \qquad \text{(由归纳法假设)}$$

结论由下列引理可得。

引理 8.6.7

设 X_1，X_2，…是一列非负独立同分布的随机变量，Y_1，Y_2，…也类似。设 N 和 M 是整值非负随机变量，它们与诸 X_i 和诸 Y_i 独立。则

$$X_i \underset{v}{\geqslant} Y_i, N \underset{v}{\geqslant} M \ \Rightarrow\ \sum_{i=1}^{N} X_i \underset{v}{\geqslant} \sum_{i=1}^{M} Y_i$$

证明 我们先证

(8.6.6)
$$\sum_{i=1}^{N} X_i \underset{v}{\geqslant} \sum_{i=1}^{M} X_i$$

以 h 记一个递增凸函数。为了证明(8.6.6)，我们必须证明

(8.6.7)
$$E\left[h\left(\sum_{i=1}^{N} X_i\right)\right] \geqslant E\left[h\left(\sum_{i=1}^{M} X_i\right)\right]$$

因为 $N \underset{v}{\geqslant} M$，且它们与诸 X_i 独立，所以如果我们能证明定义为

$$g(n) = E[h(X_1 + \cdots + X_n)]$$

的函数 $g(n)$是 n 的递增凸函数，就得不等式(8.6.7)。由于 h 递增，每个 X_i 非负，所以 $g(n)$显然是递增的。剩下证明 g 是凸的，或等价地

(8.6.8)
$$g(n+1) - g(n) \quad \text{关于 } n \text{ 是递增的}$$

为了证明这一点，令 $S_n = \sum_{i=1}^{n} X_i$，且注意到

$$g(n+1) - g(n) = E[h(S_n + X_{n+1}) - h(S_n)]$$

现在

$$\begin{aligned} &E[h(S_n+X_{n+1})-h(S_n) \mid S_n=t] \\ &\quad =E[h(t+X_{n+1})-h(t)]=f(t) \quad \text{(记作 } f\text{)} \end{aligned}$$

因为 h 是凸的，故 f 关于 t 递增。又因为 S_n 关于 n 递增，也可见 $E[f(S_n)]$关于 n 递增。然而

$$E[f(S_n)] = g(n+1) - g(n)$$

故(8.6.8)和(8.6.7)得以满足。

于是我们证明了

$$\sum_{i=1}^{N} X_i \underset{v}{\geqslant} \sum_{i=1}^{M} X_i$$

为完成证明，只需证明

$$\sum_{i=1}^{M} X_i \underset{v}{\geqslant} \sum_{i=1}^{M} Y_i$$

或等价地，对递增凸函数 h

$$E\left[h\left(\sum_{i=1}^{M} X_i\right)\right] \geqslant E\left[h\left(\sum_{i=1}^{M} Y_i\right)\right]$$

然而

$$E\left[h\left(\sum_{i=1}^{M} X_i\right)\bigg| M=m\right] = E\left[h\left(\sum_{i=1}^{m} X_i\right)\right] \quad \text{（由独立性）}$$

$$\geqslant E\left[h\left(\sum_{i=1}^{m} Y_i\right)\right] \quad \left(\text{因为} \sum_{i=1}^{m} X_i \underset{v}{\geqslant} \sum_{i=1}^{m} Y_i\right)$$

$$= E\left[h\left(\sum_{i=1}^{M} Y_i\right)\bigg| M=m\right]$$

上式两边取期望就得结论。

这样我们就证明了定理8.6.6，它说第一个过程中第 n 代的人口总量比第二个过程中的更多变。在本节的结尾我们将证明，若第二个（较少变的）过程具有与第一个过程相同的每个个体的平均后代数，则在每一代它灭绝的可能性较小。

系 8.6.8

设 μ_1 和 μ_2 分别为后代数分布 F_1 和 F_2 的均值。若 $Z_0^{(i)}=1, i=1,2, \mu_1=\mu_2=\mu, F_1 \underset{v}{\geqslant} F_2$，则对一切 n

$$P\{Z_n^{(1)}=0\} \geqslant P\{Z_n^{(2)}=0\}$$

证明 由定理8.6.6得 $Z_n^{(1)} \underset{v}{\geqslant} Z_n^{(2)}$，从而由命题8.5.1得

$$\sum_{i=2}^{\infty} P\{Z_n^{(1)} \geqslant i\} \geqslant \sum_{i=2}^{\infty} P\{Z_n^{(2)} \geqslant i\}$$

或等价地，由于

$$E[Z_n] = \sum_{i=1}^{\infty} P\{Z_n \geqslant i\} = \mu^n$$

我们有

$$\mu^n - P\{Z_n^{(1)} \geqslant 1\} \geqslant \mu^n - P\{Z_n^{(2)} \geqslant 1\}$$

或

$$P\{Z_n^{(2)} \geqslant 1\} \geqslant P\{Z_n^{(1)} \geqslant 1\}$$

这就证明了结论。

习 题

1. 若 $X \underset{\text{st}}{\geqslant} Y$，证明：$X^+ \underset{\text{st}}{\geqslant} Y^+$ 和 $Y^- \underset{\text{st}}{\geqslant} X^-$。
2. 设 $X_i \underset{\text{st}}{\geqslant} Y_i$，$i=1$，2。通过反例说明 $X_1+X_2 \underset{\text{st}}{\geqslant} Y_1+Y_2$，不一定正确。
3. (a) 若 $X \underset{\text{st}}{\geqslant} Y$，假设 X 与 Y 独立，证明 $P\{X \geqslant Y\} \geqslant \frac{1}{2}$。

   (b) 若 $P\{X \geqslant Y\} \geqslant \frac{1}{2}$，$P\{Y \geqslant Z\} \geqslant \frac{1}{2}$，且 X，Y，Z 独立，是否可以推出 $P\{X \geqslant Z\} \geqslant \frac{1}{2}$？证明或给出一个反例。
4. n 个元素中的一个被需求，以概率 P_i 需求元素 i，$i=1$，2，…，n。如果把这些元素排成一个有序名册，求使所需求的元素的位置随机极小的排法。
5. 称一个随机变量服从一个 Γ-分布，如果对某 $\lambda>0$，$\alpha>0$，它有密度函数为

   $$f(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^{\alpha-1}}{\Gamma(\alpha)}, \qquad t \geqslant 0$$

   其中

   $$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx$$

   证明：$\alpha \geqslant 1$ 时它为 IFR，$\alpha \leqslant 1$ 时为 DFR。
6. 若 X_i，$i=1,2,\cdots,n$，是独立的 IFR 随机变量，证明 $\min(X_1,\cdots,X_n)$ 也是 IFR。给出一个反例说明 $\max(X_1,\cdots,X_n)$ 不一定是。
7. 能够证明下述的关于 IFR 分布的定理。

   定理：若 F 和 G 是 IFR，则 F 和 G 的卷积 $F*G$ 也是。

   证明：用 DFR 代替 IFR 时，上述定理不真。
8. 称一个取非负整数值的随机变量为离散的 IFR，如果

   $$P\{X=k \mid X \geqslant k\} \text{关于 } k \text{ 非降}, \qquad k=0,1,\cdots$$

   证明：(a) 二项分布随机变量，(b) 泊松随机变量，(c) 负二项分布随机变量，都是离散的 IFR 随机变量。

   （提示：证明会更容易些，如果利用 IFR 分布的卷积（习题7中叙述的定理，它对离散的 IFR 分布仍成立）和极限仍然为 IFR。）

9. 证明：一个二项（n，p）分布 $B_{n,p}$，既随 n 的增加，也随 p 的增加而随机增加，即 $\overline{B}_{n,p}$，关于 n 及 p 递增。

10. 考虑一个状态为0，1，…，n 的马尔可夫链，其转移概率

$$P_{0,0}=P_{n,n}=1$$

$$P_{ij}=\begin{cases}c_i\ , & \text{若 } j=i-1\\ 1-c_i, & \text{若 } j=i+k(i)\end{cases}\qquad i=1,\ \cdots,\ n-1$$

其中 $k(i)\geqslant 0$。以 f_i 记这个马尔可夫链从状态 i 出发迟早进入状态0的概率。证明：f_i 是（c_1，…，c_{n-1}）的增函数。（提示：考虑两个这样的链，一个有（c_1，…，c_{n-1}），另一个有（$\bar{c}_1$，…，$\bar{c}_{n-1}$），其中 $\bar{c}_j\geqslant c_j$。假设它们都从状态 i 出发，耦合这二个过程，使得第一个过程的下一个状态不小于第二个过程的下一个状态。然后让第一个过程游动（保持第二个过程固定），直至它的状态与第二个过程一样或处于状态 n。如果它处于相同的状态，则重复以上的步骤。）

11. 证明：均值为 μ 和方差为 σ^2 的正态分布随 μ 增加而随机增加。当 σ^2 增加时如何？

12. 考虑一个具有转移概率矩阵(P_{ij})的马尔可夫链，且设对一切 k，$\sum\limits_{j=k}^{\infty}P_{ij}$关于 i 递增。

(a) 证明：对一切递增函数 f，$\sum\limits_{j=0}^{\infty}P_{ij}f(i)$关于 i 递增。

(b) 证明：对一切 k，$\sum\limits_{j=k}^{\infty}P_{ij}^n$关于 i 递增，其中 P_{ij}^n是 n 步转移概率，$n\geqslant 2$。

13. 设$\{N_i(t),(t\geqslant 0)\}$，$i=1,2$，为两个更新过程，分别有来到间隔分布 F_1 和 F_2。以 $\lambda_i(t)$记 F_i 的失效率函数，$i=1,2$。若对一切 s，t

$$\lambda_1(t)\geqslant\lambda_2(s)$$

证明：对任何集合 $T_1,\cdots,T_n$，

$$(N_1(T_1),\cdots,N_1(T_n))\underset{\text{st}}{\geqslant}((N_2(T_1),\cdots,N_2(T_n))$$

14. 考虑一个条件泊松过程（见第二章2.6节）。也就是，设 Λ 是一个分布为 G 的非负随机变量，$\{N(t),t\geqslant 0\}$是一个计数过程，它在 $\Lambda=\lambda$ 的条件下是一个参数为 λ 的泊松过程。以 $G_{t,n}$记给定 $N(t)=n$ 时 Λ 的条件分布。

(a) 导出 $G_{t,n}$的一个表达式。

(b) $G_{t,n}$关于 n 随机递增吗？关于 t 随机递减吗？

(c) 给定 $N(t)=n$，以 $Y_{t,n}$ 记从 t 直到下一个事件为止的时间。$Y_{t,n}$ 关于 t 随机递增吗？关于 n 随机递减吗？

15. 设 X_1 和 X_2 分别具有失效率函数 $\lambda_1(t)$ 和 $\lambda_2(t)$。证明：对一切 t，$\lambda_1(t)\geqslant\lambda_2(t)$ 当且仅当对一切 $s<t$

$$\frac{P\{X_1>t\}}{P\{X_1>s\}}\leqslant\frac{P\{X_2>t\}}{P\{X_2>s\}}$$

16. 设 X 和 Y 分别具有失效率函数 $\lambda_X(t)$ 和 $\lambda_Y(t)$，且对一切 t，$\lambda_X(t)\leqslant\lambda_Y(t)$。定义随机变量 $\overline{X}$ 如下：如果 $Y=t$，那么

$$\overline{X}=\begin{cases}t, & \text{以概率 } \lambda_X(t)/\lambda_Y(t)\\ t+X_t, & \text{以概率} 1-\lambda_X(t)/\lambda_Y(t)\end{cases}$$

其中 X_t 是一个与其它变量独立的随机变量，具有分布

$$P\{X_t>s\}=\frac{P\{X>t+s\}}{P\{X>t\}}$$

证明：$\overline{X}$ 与 X 同分布。

17. 设 F 和 G 具有失效率函数 λ_F 和 λ_G。证明：对一切 t，$\lambda_F(t)\geqslant\lambda_G(t)$ 当且仅当存在独立的、连续的随机变量 Y 和 Z，使得 Y 有分布 G，$\min(Y,Z)$ 有分布 F。

18. 一族随机变量 $\{X_\theta,\theta\in[a,b]\}$ 称为一个单调似然比族，如果 $\theta_1\leqslant\theta_2$ 时 $X_{\theta_1}\underset{\mathrm{LR}}{\leqslant}X_{\theta_2}$。证明下列随机变量族具有单调似然比：

(a) X_θ 是参数为 n，θ 的二项分布随机变量，n 固定。

(b) X_θ 是均值为 θ 的泊松随机变量。

(c) X_θ 是 $(0,\theta)$ 上的均匀分布随机变量。

(d) X_θ 是参数为 $(n,1/\theta)$ 的 Γ-随机变量，n 固定。

(e) X_θ 是参数为 (θ,λ) 的 Γ-随机变量，λ 固定。

19. 考虑一个统计推断问题，其中随机变量 X 有密度或为 f 或为 g。贝叶斯方法是规定一个 f 是真实密度的先验概率 p。f 为真实密度这一假设被接受，如果 X 的值给定时，其后验概率大于其临界值。证明：如果 $f(x)/g(x)$ 关于 x 非降，这等价于 X 的观察值大于某临界值时接受 f。

20. 我们有 n 项工作，第 i 项工作需花随机时间 X_i 去做。这些工作必须相继地进行。我们的目标是使到（固定的）时刻 t 为止所进行的工作项数随机最大。

(a) 确定最优策略，如果

$$X_i \underset{\text{LR}}{\leqslant} X_{i+1}, \qquad i=1, 2, \cdots, n-1$$

(b) 如果我们仅假设

$$X_i \underset{\text{st}}{\leqslant} X_{i+1}, \qquad i=1, 2, \cdots, n-1$$

则如何?

21. 有 n 种作物，与第 i 种作物联系着一个随机变量 X_i，$i=1, 2, \cdots, n$。如果在时刻 t 将第 i 种作物种植到大田里，那么它的田间寿命为 $X_i e^{-\alpha t}$。当 $X_i \underset{\text{LR}}{\geqslant} X_{i+1}$，$i=1, 2, \cdots, n-1$时，怎样的种植顺序使全部作物的总田间寿命随机极大?注意，若 $n=2$及使用的顺序为1，2，则总的田间寿命为 $X_1+X_2 e^{-\alpha X_1}$。

22. 证明：具有下列密度的随机变量有递增似然比，即密度的对数是凹函数。

(a) Γ-分布：$f(x)=\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}/\Gamma(\alpha)$，$\quad \alpha\geqslant 1$

(b) 威布尔分布：$f(x)=\alpha\lambda(\lambda x)^{\alpha-1}e^{-(\lambda x)^{\alpha}}$，$\quad \alpha\geqslant 1$

(c) 截为正的正态分布：

$$f(x)=\frac{1}{a\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}, \qquad x>0$$

23. 设 $X_1, X_2, \cdots, X_n$ 相互独立，$Y_1, \cdots, Y_n$ 相互独立。或 $X_i \underset{\text{v}}{\geqslant} Y_i$，$i=1, 2, \cdots, n$，证明：

$$E[\max(X_1, \cdots, X_n)] \geqslant E[\max(Y_1, \cdots, Y_n)]$$

不假设独立性的给出一个反例。

24. 证明：F 是一个 NBUE 随机变量的分布当且仅当对一切 a

$$\overline{F}_e(a) \leqslant \overline{F}(a)$$

其中 F_e 是 F 的平衡分布，定义为

$$F_e(a)=\int_0^a \frac{\overline{F}(x)}{\mu}dx$$

而$\mu=\int_0^\infty \overline{F}(x)dx$。

25. 证明下述命题或给出反例：若 $F \underset{\text{v}}{\geqslant} G$，则 $N_F(t) \underset{\text{v}}{\geqslant} N_G(t)$，其中$\{N_H(t), t\geqslant 0\}$是来到间隔分布为 H 的更新过程，$H=F, G$。

26. 设 $X_1, X_2, \cdots, X_n$ 相互独立，且 $P\{X_i=1\}=P_i=1-P\{X_i=0\}$，$i=1, 2, \cdots, n$。证明：

$$\sum_{i=1}^{n} X_i \underset{\text{v}}{\leqslant} \text{Bin}(n, \overline{p})$$

其中 Bin $(n, \bar{p})$，是参数为 n 和 $\bar{p}=\sum_{i=1}^{n} p_i/n$ 的二项分布随机变量。

（提示：先对$n=2$证明结论。利用上述结果证明$\sum_{i=1}^{n} X_i \underset{v}{\leqslant} Y$，其中 Y 是均值为 $n\bar{p}$的泊松随机变量。因此，例如一个泊松随机变量比一个有相同均值的二项分布随机变量更多变。）

27. 设 $P\{X=0\}=1-M/2$，$P\{X=2\}=M/2$，其中$0<M<2$。若 Y 是一个非负整值随机变量，使得 $P\{Y=1\}=0$及 $E[Y]=M$，证明：$X \underset{v}{\leqslant} Y$。

28. 延森不等式说，对一个凸函数 f

$$E[f(X)] \geqslant f(E[X])$$

(a) 证明延森不等式。

(b) 若 X 有均值 $E[X]$，证明：

$$X \underset{v}{\geqslant} E[X]$$

其中 $E[X]$为常数随机变量。

(c) 设存在一个随机变量 Z，满足 $E[Z|Y]\geqslant 0$，且 X 与 $Y+Z$ 同分布。证明 $X \underset{v}{\geqslant} Y$。实际上，可以证明这是 $X \underset{v}{\geqslant} Y$ 的一个充要条件（另一个方向的证明是困难的）。

29. 如果 $E[X]=0$，证明：$c\geqslant 1$时，$cX \underset{v}{\geqslant} X$。

参考文献

有关随机大于方面的其它结果，读者应参看文献5、8和9。有关 IFR 和 DFR 随机变量的其它材料应参看文献1。耦合是一种有价值又有用的技巧，它起源于文献6；实际上，8.2.2节中所用的方法就来自文献6。失效率序在文献10中使用。导致并证明命题8.3.2和系8.3.3的方法取自文献2和3。有关似然比序的其它材料，读者应参看文献7。命题8.4.2来自文献4。文献5和9对变动性序是有用的。8.6.1节中的结果来自文献11；8.6.2节中的结果来自文献1；8.6.3节中的结果来自文献12。

1. R. E. Barlow, and F. Proschan, *Statistical Theory of Reliability and Life Testing*, Holt, New York, 1975.

2. M., Brown, "Bounds, Inequalities and Monotonicity Properties for Some

Specicalized Renewal Processes," *Annals of Probability*, 8(1980), pp. 227—240.

3. M., Brown, "Further Monotonicity Properties for specialized Renewal Processes," *Annals of Probability*, 9(1981), pp. 891—895.
4. M. Brown, and H. Solomon, "Optimal Issuing Policies Under Stochastic Field Lives," *Journal of Applied Probability*, 10, No. 4(1973)pp. 761—768.
5. S. L. Brumelle, and R. G. Vickson, "A Unified Approach to Stochastic Dominance," in *Stochastic Optimization Models in Finance*, edited by W. Ziemba and R. Vickson, Academic Press, New York, 1975.
6. W., Doeblin, "Sur Deux Problemes de M. Kolmogoroff Concernant Les Chaines Denombrables," *Bulletin Societe Mathematique de France*, 66 (1938)pp. 210—220。
7. E., Lehmann, *Testing Statistical Hypothesis*, Wiley, New York, 1959.
8. Lehmann, E., "Some Concepts of Dependence," *Annals of Mathematical Statistics*, 37(1966), pp. 1137—1153.
9. A. Marshall, and I. Oklin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979。
10. M. Pinedo and S. M. Ross, "Scheduling Jobs Subject to Nonhomogeneous Poisson Shocks," *Management Science*, 26, No. 12(1980), pp. 1250—1257.
11. T. Rolski and D. Stoyan, "On the Comparison of Waiting Times in GI/G/1 Queues," *Operations Research*, 24(1976), pp. 197—200.
12. S. M. Ross and Z. Schechner, "Some Reliability Applications of the Variability Ordering," Operations Research Center Report, University of California Berkeley, 1982.

习题选解与部分答案

第一章

5. (a) $P(N_1=n_1,N_2=n_2,\cdots,N_r=n_r)=(n!/\prod_{i=1}^{r}n_i!)\prod_{i=1}^{r}P_i^{n_i}$

其中 $n_i=0$，…，n 及 $\sum_{i=1}^{r}n_i=n$

(b) $E[N_i]=nP_i, E[N_i^2]=nP_i-nP_i^2+n^2P_i^2$

$E[N_j]=nP_j, E[N_j^2]=nP_j-nP_j^2+n^2P_j^2$

$E[N_iN_j]=E[E[N_iN_j|N_j]]$

$$E[N_iN_j|N_j=m]=mE[N_i|N_j=m]$$
$$=m(n-m)\frac{P_i}{1-P_j}=\frac{nmP_i-m^2P_i}{1-P_j}$$

$$E[N_iN_j]=\frac{nE[N_j]P_i-E[N_j^2]P_i}{1-P_j}=\frac{n^2P_jP_i-nP_iP_j+nP_j^2P_i-n^2P_jP_i}{1-P_j}$$
$$=\frac{n^2P_jP_i(1-P_j)-nP_iP_j(1-P_j)}{1-P_j}=n^2P_iP_j-nP_iP_j$$

$\mathrm{Cov}(N_i,N_j)=E[N_iN_j]-E[N_i]E[N_j]=-nP_iP_j$， $i\neq j$

(c) 令 $I_j=\begin{cases}1, & \text{若结果 } j \text{ 从未出现}\\ 0, & \text{其它}\end{cases}$

$E[I_j]=(1-P_j)^n$， $\mathrm{Var}[I_j]=(1-P_j)^n(1-(1-P_j)^n)$

$E[I_iI_j]=(1-P_i-P_j)^n$， $i\neq j$

未出现的结果的个数$=\sum_{j=1}^{r}I_j$

$$E[\sum_{j=1}^{r}I_j]=\sum_{j=1}^{r}(1-P_j)^n$$

$$\mathrm{Var}[\sum_{j=1}^{r}I_j]=\sum_{j=1}^{r}\mathrm{Var}[I_j]+\sum_{i\neq j}\sum\mathrm{Cov}(I_i,I_j)$$
$$\mathrm{Cov}(I_i,I_j)=E[I_iI_j]-E[I_i]E[I_j]$$

$$= (1 - P_i - P_j)^n - (1 - P_i)^n (1 - P_j)^n$$

$$\mathrm{Var}[\sum_{j=1} I_j] = \sum_{j=1}^{r} (1 - P_j)^n (1 - (1 - P_j)^n)$$
$$+ \sum_{i \neq j}\sum [(1 - P_i - P_j)^n - (1 - P_i)^n (1 - P_j)^n]$$

6. (a) 令 $$I_j = \begin{cases} 1, & \text{若在时刻 } j \text{ 有一个纪录} \\ 0, & \text{其它} \end{cases}$$

$$N_n = \sum_{j=1}^{n} I_j$$

$$E[N_n] = \sum_{j=1}^{n} E[I_j] = \sum_{j=1}^{n} \frac{1}{j}$$

$$\mathrm{Var}[N_n] = \sum_{j=1}^{n} \mathrm{Var}[I_j] = \sum_{j=1}^{n} \frac{1}{j}(1 - \frac{1}{j})$$,因为 I_j 相互独立。

(b) 令 $T = \min\{n: n > 1$ 且在时刻 n 产生一个纪录$\}$

$T > n \iff X_1 = X_1, X_2, \cdots, X_n$ 中的最大者

$$E[T] = \sum_{n=1}^{\infty} P\{T > n\} = \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

$$P\{T = \infty\} = \lim_{n \to \infty} P\{T > n\} = 0$$

(c) 以 T_y 记第一个大于 y 的纪录值的时刻，以 X_{T_y} 记时刻 T_y 的纪录值。

$$P\{X_{T_y} > x | T_y = n\} = P\{X_n > x | X_1 < y, X_2 < y, \cdots, X_{n-1} < y, X_n > y\}$$
$$= P\{X_n > x | X_n > y\}$$
$$= \begin{cases} 1 & , \quad x < y \\ \overline{F}(x)/\overline{F}(y), & \quad x > y \end{cases}$$

由于 $P\{X_{T_y} > x | T_y = n\}$ 不依赖于 n，得出 T_y 与 X_{T_y} 独立的结论。

12. $P\{N = 0\} = P\{U_1 < e^{-\lambda}\} = e^{-\lambda}$。假设

$$P\{N = n\} = P\{U_1 \geqslant e^{-\lambda}, U_1U_2 \geqslant e^{-\lambda}, \cdots, U_1 \cdots U_n \geqslant e^{-\lambda}, U_1 \cdots U_{n+1} < e^{-\lambda}\}$$
$$= e^{-\lambda}\lambda^n / n!$$

从而

$$P\{N = n + 1\}$$
$$= \int_0^1 P\{U_1 \geqslant e^{-\lambda}, \cdots, U_1 \cdots U_{n+1} \geqslant e^{-\lambda}, U_1 \cdots U_{n+2} < e^{-\lambda} | U_1 = x\} dx$$
$$= \int_{e^{-\lambda}}^1 P\left\{U_2 \geqslant \frac{e^{-\lambda}}{x}, \cdots, U_2 \cdots U_{n+1} \geqslant \frac{e^{-\lambda}}{x}, U_2 \cdots U_{n+2} < \frac{e^{-\lambda}}{x}\right\} dx$$
$$= \int_{e^{-\lambda}}^1 e^{-(\lambda + \log x)} \frac{(\lambda + \log x)^n}{n!} dx$$

其中最后的等式得自归纳法假设，因为$e^{-\lambda}/x=e^{-(\lambda+\log x)}$。由上式我们有

$$P\{N=n+1\}=\frac{1}{n!}\int_{e^{-\lambda}}^{1}\frac{e^{-\lambda}(\lambda+\log x)^n}{x}dx$$

$$=\frac{e^{-\lambda}}{n!}\int_0^{\lambda}y^n dy \qquad (\text{由 } y=\lambda+\log x)$$

$$=e^{-\lambda}\lambda^{n+1}/(n+1)!$$

这就完成了归纳法证明。

13. $\mathrm{Var}(X|Y)=E[(X-E(X|Y))^2|Y]$

$$=E[X^2-2XE[X|Y]+(E[X|Y])^2|Y]$$

$$=E[X^2|Y]-2(E[X|Y])^2+(E[X|Y])^2$$

$$=E[X^2|Y]-(E[X|Y])^2$$

$$\mathrm{Var}(X)=E[X^2]-(E[X])^2$$

$$=E[E[X^2|Y]]-(E[E[X|Y]])^2$$

$$=E[\mathrm{Var}(X|Y)+(E[X|Y])^2]-(E[E[X|Y]])^2$$

$$=E[\mathrm{Var}(X|Y)]+E[(E[X|Y])^2]-(E[E[X|Y]])^2$$

$$=E[\mathrm{Var}(X|Y)]+\mathrm{Var}(E[X|Y])$$

22. $P\{X_1<X_2|\min(X_1,X_2)=t\}=\dfrac{P\{X_1<X_2,\min(X_1,X_2)=t\}}{P\{\min(X_1,X_2)=t\}}$

$$=\frac{P\{X_1=t,X_2>t\}}{P\{X_1=t,X_2>t\}+P\{X_2=t,X_1>t\}}$$

$$=\frac{P\{X_1=t\}P\{X_2>t\}}{P\{X_1=t\}P\{X_2>t\}+P\{X_2=t\}P\{X_1>t\}}$$

$$P\{X_2=t\}=\lambda_2(t)P\{X_2>t\}$$

$$P\{X_1>t\}=P\{X_1=t\}/\lambda_1(t)$$

$$P\{X_2=t\}P\{X_1>t\}=\frac{\lambda_2(t)}{\lambda_1(t)}P\{X_1=t\}P\{X_2>t\}$$

从而

$$P\{X_1<X_2|\min(X_1,X_2)=t\}=\frac{1}{1+\dfrac{\lambda_2(t)}{\lambda_1(t)}}$$

$$=\frac{\lambda_1(t)}{\lambda_1(t)+\lambda_2(t)}$$

第二章

6. 以 N 记失效的部件数，要求的答案是 $E[N]/(\mu_1+\mu_2)$，其中

$$E[N]=\sum_{k=\min(n,m)}^{n+m-1}k\left[\binom{k-1}{n-1}\left(\frac{\mu_1}{\mu_1+\mu_2}\right)^n\left(\frac{\mu_2}{\mu_1+\mu_2}\right)^{k-n}\right.$$
$$\left.+\binom{k-1}{m-1}\left(\frac{\mu_1}{\mu_1+\mu_2}\right)^{k-m}\left(\frac{\mu_2}{\mu_1+\mu_2}\right)^m\right]$$

其中 $i>k-1$ 时 $\binom{k-1}{i}=0$。

7. 由于 $(S_1,S_2,S_3)=(s_1,s_2,s_3)$ 等价于 $(X_1,X_2,X_3)=(s_1,s_2-s_1,s_3-s_2)$，故得 S_1,S_2,S_3 的联合密度为

$$f(s_1,s_2,s_3)=\lambda e^{-\lambda s_1}\lambda e^{-\lambda(S_2-S_1)}\lambda e^{-\lambda(s_3-s_2)}$$
$$=\lambda^3 e^{-\lambda s_3},\quad 0<s_1<s_2<s_3$$

8. (a)
$$P\left\{\frac{-\log U_i}{\lambda}\leqslant x\right\}=P\left\{\log\left(\frac{1}{U_i}\right)\leqslant\lambda x\right\}$$
$$=P\{1/U_i\leqslant e^{\lambda x}\}$$
$$=P\{U_i\geqslant e^{-\lambda x}\}$$
$$=1-e^{-\lambda x}$$

(b) 以 X_i 记泊松过程的来到间隔时间，则 $N(1)$（到时刻1的事件数）将等于使下式成立的 n

$$\sum_{i=1}^{n}X_i<1<\sum_{i=1}^{n+1}X_i$$

或等价地，使下式成立的 n

$$-\sum_{i=1}^{n}\log U_i<\lambda<-\sum_{i=1}^{n+1}\log U_i$$

或等价地

$$\sum_{i=1}^{n}\log U_i\geqslant-\lambda>\sum_{i=1}^{n+1}\log U_i$$

或

$$\prod_{i=1}^{n}U_i\geqslant e^{-\lambda}>\prod_{i=1}^{n+1}U_i$$

由于 $N(1)$ 是均值为1的泊松随机变量，结论得证。

9. 以 N_{ij} 记在第 i 层乘上电梯，在第 j 层离去的人数，则 N_{ij} 是均值为 $\lambda_i P_{ij}$ 的泊松变量，且全部 N_{ij}（$i\geqslant0$，$j\geqslant i$）相互独立。因此：

(a) $E[0_j]=E[\sum_i N_{ij}]=\sum_i\lambda_i P_{ij}$

(b) $0_j=\sum_i N_{ij}$ 是均值为 $\sum_i\lambda_i P_{ij}$ 的泊松变量

(c) 0_i 与 0_k 独立

10. (a) N_i 是服从负二项分布的变量。即

$$P\{N_i = k\} = \binom{k-1}{n_i - 1} P_i^{n_i} (1 - P_i)^{k-n_i}, \qquad k \geqslant n_i$$

(b) 否

(c) T_i 服从参数为 n_i 和 P_i 的 Γ-分布

(d) 是

(e) $$\begin{aligned} E[T] &= \int_0^\infty P\{T > t\} dt \\ &= \int_0^\infty P\{T_i > t, i = 1, \cdots, r\} dt \\ &= \int_0^\infty \left(\prod_{i=1}^r P\{T_i > t\}\right) dt \quad \text{(由独立性)} \\ &= \int_0^\infty \left(\prod_{i=1}^r \int_t^\infty \frac{P_i e^{-P_i x} (P_i x)^{n_i - 1}}{(n_i - 1)!} dx\right) dt \end{aligned}$$

(f) $T = \sum_{i=1}^N X_i$，其中 X_i 是第 $i-1$ 次与第 i 次投掷之间的时间。由于 N 与 X_i 的序列独立，我们得到

$$\begin{aligned} E[T] &= E[E[T|N]] \\ &= E[NE[X]] \\ &= E[N] \quad \text{(因为 } E[X] = 1\text{)} \end{aligned}$$

17. 称一辆进入的汽车是1-型的，若它在时刻 t 处于 a 与 b 之间。因此一辆在时刻 s $(s<t)$，进入的汽车是1-型的，若其速度 V 使得 $a<(t-s)V<b$，故它为1-型的概率为

$$F\left(\frac{b}{t-s}\right) - F\left(\frac{a}{t-s}\right)$$

因此1-型汽车的个数是泊松变量，均值为

$$\lambda\int_0^t \left(F\left(\frac{b}{t-s}\right) - F\left(\frac{a}{t-s}\right)\right) ds$$

18. 以与定理2.3.1同样的方式能证明，给定 $S_n = t$ 时，$S_1, \cdots, S_{n-1}$ 的分布等同于一组 $n-1$ 个独立的 $(0,t)$ 上均匀分布的随机变量的顺序统计量的分布。第二种方法较直观些是

$$\begin{aligned} S_1, \cdots, S_{n-1} | S_n = t &= S_1, \cdots, S_{n-1} | N(t^-) = n-1, N(t) = n \\ &= S_1, \cdots, S_{n-1} | N(t^-) = n-1 \qquad \text{(由独立增量)} \\ &= S_1, \cdots, S_{n-1} | N(t) = n-1 \end{aligned}$$

28. (a)它不可能有独立增量,因为知道任一区间中的事件数将改变 Λ 的分布。

(b)知道$\{N(s),0\leqslant s\leqslant t\}$等价于知道 $N(t)$ 及来到时刻 $S_1,\cdots,S_{N(t)}$。现在(直观地论证)对$0<s_1<\cdots<s_n<t$,

$$P\{\Lambda=\lambda,N(t)=n,S_1=s_1,\cdots,S_n=s_n\}$$
$$=P\{\Lambda=\lambda\}P\{N(t)=n\mid\Lambda=\lambda\}P\{S_1=s_1,\cdots,S_n=s_n\mid\Lambda=t,N(t)=n\}$$
$$=dG(t)e^{-\lambda t}\frac{(\lambda t)^n}{n!}\cdot\frac{n!}{t^n}\qquad\text{(由定理 2.3.1)}$$

从而

$$P\{\Lambda\in(\lambda,\lambda+d\lambda)\mid N(t)=n,S_1=s_1,\cdots,S_n=s_n\}$$
$$=\frac{e^{-\lambda t}(\lambda t)^n dG(\lambda)}{\int_0^\infty e^{-\lambda t}(\lambda t)^n dG(\lambda)}$$

因此 Λ 的条件分布只依赖于 $N(t)$。这是因为给定 $N(t)$的值,不管 Λ 的值是什么,$S_1,\cdots,S_{N(t)}$的分布与来自$(0,t)$上均匀分布的母体的顺序统计量的分布相同。

(c) $P\{t$ 之后第一个事件的时刻大于 $t+s\mid N(t)=n\}$

$$=\frac{\int_0^\infty e^{-\lambda s}e^{-\lambda t}(\lambda t)^n dG(\lambda)}{\int_0^\infty e^{-\lambda t}(\lambda t)^n dG(\lambda)}$$

(d) $$\lim_{h\to0}\int_0^\infty\frac{1-e^{-\lambda h}}{h}dG(\lambda)=\int_0^\infty\lim_{h\to0}\left(\frac{1-e^{-\lambda h}}{h}\right)dG(\lambda)$$
$$=\int_0^\infty\lambda dG(\lambda)$$

(e) 同分布但不独立。

29. (a) $$P\{N(t)=n\}=\int_0^\infty e^{-\lambda t}\frac{(\lambda t)^n}{n!}\alpha e^{-\alpha\lambda}\frac{(\alpha\lambda)^{m-1}}{(m-1)!}d\lambda$$
$$=\frac{\alpha^m t^n(m+n-1)!}{n!(m-1)!(\alpha+t)^{m+n}}\int_0^\infty(\alpha+t)e^{-(\alpha+t)\lambda}\frac{[(\alpha+t)\lambda]^{m+n-1}}{(m+n-1)!}d\lambda$$
$$=\binom{m+n-1}{n}\left(\frac{\alpha}{\alpha+t}\right)^m\left(\frac{t}{\alpha+t}\right)^n$$

(b) $$P\{\Lambda=\lambda\mid N(t)=n\}=\frac{P\{N(t)=n\mid\Lambda=\lambda\}P\{\Lambda=\lambda\}}{P\{N(t)=n\}}$$
$$=\frac{e^{-\lambda t}\frac{(\lambda t)^n}{n!}\alpha e^{-\alpha\lambda}\frac{(\alpha\lambda)^{m-1}}{(m-1)!}}{\binom{m+n-1}{n}\left(\frac{\alpha}{\alpha+t}\right)^m\left(\frac{t}{\alpha+t}\right)^n}$$

$$= (\alpha + t)e^{-(\alpha+t)\lambda}\frac{[(\alpha + t)\lambda]^{m+n-1}}{(m + n - 1)!}$$

(c) 由习题26(d)，答案是条件分布的均值，由(b)它等于$(m+n)/(\alpha+t)$。

第三章

14. $g = h + g * F$

$$= h + (h + g * F) * F = h + h * F + g * F_2$$

$$= h + h * F + (h + g * F) * F_2$$

$$= h + h * F + h * F_2 + g * F_3$$

$$\vdots$$

$$= h + h * F + h * F_2 + \cdots + h * F_n + g * F_{n+1}$$

令$n \to \infty$并利用$F_n \to 0$得

$$g = h + h * \sum_{n=1}^{\infty} F_n = h + h * m$$

(a) $$P(t) = \int_0^{\infty} P\{\text{在时刻}t\text{开} \mid Z_1 + Y_1 = s\}dF(s)$$

$$= \int_0^t P(t - s)dF(s) + \int_t^{\infty} P\{Z_1 > t \mid Z_1 + Y_1 = s\}dF(s)$$

$$= \int_0^t P(t - s)dF(s) + P\{Z_1 > t\}$$

(b) $$g(t) = \int_0^{\infty} E[A(t) \mid X_1 = s]dF(s)$$

$$= \int_0^t g(t - s)dF(s) + \int_t^{\infty} tdF(s)$$

$$= \int_0^t g(t - s)dF(s) + t\overline{F}(t)$$

(c) $$P(t) \to \frac{\int_0^{\infty} P\{Z_1 > t\}dt}{\mu_F} = \frac{E[Z]}{E[Z] + E[Y]}$$

$$g(t) \to \frac{\int_0^{\infty} t\overline{F}(t)dt}{\mu} = \frac{\int_0^{\infty} t\int_t^{\infty} dF(s)dt}{\mu}$$

$$= \frac{\int_0^{\infty}\int_0^s tdtdF(s)}{\mu} = \frac{\int_0^{\infty} s^2 dF(s)}{2\mu} = \frac{E[X^2]}{2E[X]}$$

20. $$E[R_{N(t)+1}] = \int_0^t E[R_{N(t)+1} \mid S_{N(t)} = s]\overline{F}(t - s)dm(s)$$

$$
\begin{aligned}
&+ E[R_{N(t)+1} | S_{N(t)} = 0]\overline{F}(t) \\
&= \int_0^t E[R_1 | X_1 > t - s]\overline{F}(t-s)dm(s) \\
&\quad + E[R_1 | X_1 > t]\overline{F}(t) \\
&\to \int_0^\infty E[R_1 | X_1 > t]\overline{F}(t)dt/\mu \\
&= \int_0^\infty \int_t^\infty E[R_1 | X_1 = s]dF(s)dt/\mu \\
&= \int_0^\infty \int_0^s dtE[R_1 | X_1 = s]dF(s)/\mu \\
&= \int_0^\infty sE[R_1 | X_1 = s]dF(s)/\mu \\
&= E[R_1 X_1]/\mu
\end{aligned}
$$

其中 $\mu = E[X_1]$。我们假设了 $E[R_1X_1] < \infty$，这蕴含 $t \to \infty$ 时

$$E[R_1 | X_1 > t]\overline{F}(t) \to 0$$

由于 $\mathrm{Var}(X) > 0$，$E[X^2] > E^2[X]$，除非 X 以概率1为常数。

24. (a) 两个过程都是再生的。

(b) 若 T 是一个循环的时间，N 是受到服务的顾客数，则

$$V = E[\int_0^T V(s)ds]/E[T], \quad W_Q = \frac{E[D_1 + \cdots + D_N]}{E[N]}$$

设想在任何时刻每个人以等于其剩余服务时间的速率支付。则

$$一个循环中的酬劳 = \int_0^T V(s)ds$$

又以 Y_i 记第 i 个顾客的服务时间，

$$
\begin{aligned}
一个循环中的酬劳 &= \sum_{i=1}^N [D_i Y_i + \int_0^{Y_i}(Y_i - t)dt] \\
&= \sum_{i=1}^N D_i Y_i + \sum_{i=1}^N \frac{Y_i^2}{2}
\end{aligned}
$$

因此，

$$
\begin{aligned}
E[一个循环中的酬劳] &= E[\sum_{i=1}^N D_i Y_i] + E[\sum_{i=1}^N \frac{Y_i^2}{2}] \\
&= E[\sum_{i=1}^N D_i Y_i] + \frac{E[N]E[Y^2]}{2} \quad (由瓦尔德等式)
\end{aligned}
$$

现在

$$\frac{E[\sum_{i=1}^N D_i Y_i]}{E[N]} = \lim_{n\to\infty} \frac{E[D_1Y_1 + \cdots + D_nY_n]}{n}$$

$$= E[Y]\lim_{n\to\infty}\frac{E[D_1+\cdots+D_n]}{n} \quad (\text{由于 } D_i \text{ 与 } Y_i \text{ 独立})$$
$$= E[Y]W_Q$$

所以从上得到

$$E[\text{一个循环中的酬劳}] = E[N]\left(E[Y]W_Q + \frac{E[Y^2]}{2}\right)$$

令此式与 $E[\int_0^T V(s)ds] = VE[T]$ 相等并利用关系式 $E[T] = E[N]/\lambda$ 就证得恒等式。

25. 设 $P\{X<Y\}=1$。例如,$P\{X=1\}=1$,Y 在$(2,3)$上均匀分布及 $k=3$。

26. (a) 再生过程。

(b) $E[\text{一个循环中 } i \text{ 个包裹在等待的时间}]/\mu_F$,其中

$$E[i\text{ 个包裹在等待的时间}] = \int_0^\infty E[\text{时间}\mid\text{循环长为 } x]dF(x)$$
$$= \int_0^\infty \sum_{j=i}^\infty E[\text{时间}\mid\text{长为 } x, N(x)=j]e^{-\lambda x}\frac{(\lambda x)^j}{j!}dF(x)$$
$$= \int_0^\infty \sum_{j=i}^\infty \frac{x}{j+1}e^{-\lambda x}\frac{(\lambda x)^j}{j!}dF(x)$$

最后的等式得自第二章习题12。

27. $\lim_{t\to\infty}\int_0^t r(X(s))ds/t = E[\int_0^T r(X(s))ds]/E[T]$

$$= E[\sum_j r(j)(\text{在 } T \text{ 中处于 } j \text{ 的时间})]/E[T]$$
$$= \sum_j r(j)P_j$$

第四章

10. 设 $i\leftrightarrow j$, i 是正常返的。设 m 使 $P_{ij}^m>0$。以 N_k 记链第 k 次处于状态 i 的时间,令

$$I_k = \begin{cases} 1, & \text{若 } X_{N_k+m}=j \\ 0, & \text{其它} \end{cases}$$

由强大数定律

$$\sum_{k=1}^n \frac{I_k}{n} \to P_{ij}^m > 0$$

因此

$$\lim_{n\to\infty}\frac{\text{到时刻 } N_n+m \text{ 到达 } j \text{ 的次数}}{n}\frac{n}{N_n+m} \geqslant P_{ij}^m\frac{1}{E[T_{ii}]} > 0$$

其中 T_{ii} 是相继到达状态 i 之间的时间。因此 j 也是正常返的。若 i 是零常返的，则由于 $i\leftrightarrow j$，常返性是类性质，故 j 是常返的。若 j 是正常返，由上所证 i 也是。因此 j 是零常返的。

11. 设 i 是零常返的，以 C 记与 i 相通的状态组成的类，C 中全部状态都是零常返的，这蕴含了对一切 $j\in C$

$$\lim_{n\to\infty} P_{ij}^n = 0$$

但这是一个矛盾，因为 $\sum_{j\in C} P_{ij}^n = 1$，$C$ 是一个有限集。同样的理由，有限链的状态不可能都是非常返的。

13. (a) 令状态为他身边所有的伞数。转移概率为

$$P_{0,r}=1,\quad P_{i,r-i}=1-p,\quad P_{i,r-i+1}=p,\quad i=1,\cdots,r$$

(b) 极限概率的方程为

$$\pi_r = \pi_0 + \pi_1 p$$

$$\pi_j = \pi_{r-j}(1-p) + \pi_{r-j+1} p,\ j=1,\cdots,r-1$$

$$\pi_0 = \pi_r(1-p)$$

容易验证它们被

$$\pi_i = \begin{cases} q/(r+q), & 若\ i=0 \\ 1/(r+q), & 若\ i=1,\cdots,r \end{cases}$$

所满足，其中 $q=1-p$

(c) $p\pi_0 = pq/(r+q)$

14. 令 $I_n^j = \begin{cases} 1, & 若\ X_n=j \\ 0, & 其它 \end{cases}$ 则

$$\begin{aligned}
\pi_j &= \lim_{n\to\infty} E\Big[\sum_{k=1}^n I_k^j\Big]/n \\
&= \lim_{n\to\infty} E\Big[\sum_{k=1}^n \sum_i I_{k-1}^i I_k^j\Big]/n \\
&= \lim_{n\to\infty} \sum_{k=1}^n \sum_i E[I_{k-1}^i I_k^j]/n \\
&= \lim_{n\to\infty} \sum_{k=1}^n \sum_i E[I_{k-1}^i] P_{ij}/n \\
&= \lim_{n\to\infty} \sum_i P_{ij} \sum_{k=1}^n E[I_{k-1}^i]/n \\
&= \sum_i P_{ij} \lim_{n\to\infty} E\Big[\sum_{k=1}^n I_{k-1}^i\Big]/n
\end{aligned}$$

$$=\sum_i \pi_i P_{ij}$$

15. (a) 对 i 之后到达的状态取条件。

(b) $P\{T_i>n\}=\sum_j P\{T_i>n \mid \text{下一个状态是 } j\}P_{ij}$

$$=\sum_{j\neq 0} P\{T_j>n-1\}P_{ij}$$

16. 这马尔可夫链为正常返的当且仅当方程组

$$y_0=y_1q_1$$

$$y_j=y_{j+1}q_{j+1}+y_{j-1}p_{j-1},\quad j\geqslant 1$$

有一组解 $y_j\geqslant 0$，$\sum_j y_j=1$。这些方程可重写为

$$y_0=y_1q_1$$

$$y_{j+1}q_{j+1}-y_jp_j=y_jq_j-y_{j-1}p_{j-1},\quad j\geqslant 1$$

由此可得

$$y_{j+1}q_{j+1}=y_jp_j,\qquad j\geqslant 0$$

因此，
$$y_{j+1}=y_0\frac{p_0\cdots p_j}{q_1\cdots q_{j+1}},\qquad j\geqslant 0$$

所以，随机游动为正常返的充要条件是

$$\sum_{j=0}^{\infty}\frac{p_0\cdots p_j}{q_1\cdots q_{j+1}}<\infty$$

20. 由于 $P\{X_i-Y_i=1\}=P_1(1-P_2)$，$P\{X_i-Y_i=-1\}=(1-P_1)P_2$，如果我们仅在 $X_i-Y_i\neq 0$ 时观察 (X_i,Y_i)，那么观察到的就是一个简单随机游动，其

$$P=\frac{P_1(1-P_2)}{P_1(1-P_2)+P_2(1-P_1)}$$

因此，由赌徒输光问题的结果，

$$P\{\text{错误}\}=P\{\text{在上升 } M \text{ 之前下降 } M\}$$

$$=1-\frac{1-(1-(q/p)^M)}{1-(q/p)^{2M}}=\frac{(q/p)^M(1-(q/p)^M}{1-(q/p)^{2M}}$$

$$=\frac{(q/p)^M}{1+(q/p)^M}=\frac{1}{(p/q)^M+1}=\frac{1}{1+\lambda^M}$$

由瓦尔德等式

$$E\left[\sum_{i=1}^{N}(X_i-Y_i)\right]=E[N](P_1-P_2)$$

或

$$E[N](P_1-P_2)=M\frac{\lambda^M}{1+\lambda^M}-\frac{M}{1+\lambda^M}=\frac{M(\lambda^M-1)}{1+\lambda^M}$$

27. 对1，2，…，n 的任一置换 i_1，i_2，…，i_n，以 $\pi(i_1,i_2,\cdots,i_n)$ 记在移前一位规则下的极限概率。由时间可逆性，对一切置换我们有

$$(*)P_{i_{j+1}}\pi(i_1,\cdots,i_j,i_{j+1},\cdots,i_n)=P_{i_j}\pi(i_1,\cdots,i_{j+1},i_j,\cdots,i_n)$$

现在所要取出的元素的平均位置可表为

$$\begin{aligned}
\text{平均位置}&=\sum_i P_iE[\text{元素 } i \text{ 的位置}]\\
&=\sum_i P_i[1+\sum_{j\neq i}P\{\text{元素 } j \text{ 在元素 } i \text{ 之前}\}]\\
&=1+\sum_i\sum_{j\neq i}P_iP\{e_j \text{ 在 } e_i \text{ 之前}\}\\
&=1+\sum_{i<j}[P_iP\{e_j \text{ 在 } e_i \text{ 之前}\}+P_jP\{e_i \text{ 在 } e_j \text{ 之前}\}]\\
&=1+\sum_{i<j}[P_iP\{e_j \text{ 在 } e_i \text{ 之前}\}+P_j(1-P\{e_j \text{ 在 } e_i \text{ 之前}\})]\\
&=1+\sum_{i<j}\sum(P_i-P_j)P\{e_j \text{ 在 } e_i \text{ 之前}\}+\sum_{i<j}\sum P_j
\end{aligned}$$

因此，为使所要取出的元素的平均位置最小，我们要使 $P\{e_j$ 在 e_i 之前$\}$ 在 $P_j>P_i$ 时尽可能地大，在 $P_i>P_j$ 时尽可能地小，对于移到最前面的规则我们有

$$P\{e_j \text{ 在 } e_i \text{ 之前}\}=\frac{P_j}{P_j+P_i}$$

因为在移到最前面的规则之下，元素 j 在元素 i 之前当且仅当最后一次的取出 i 或 j 的要求是要取出 j。所以为了证明移前一位规则比移到最前面的规则好，只要证明在移前一位规则之下，当 $P_j>P_i$ 时

$$P\{e_j \text{ 在 } e_i \text{ 之前}\}>\frac{P_j}{P_j+P_i}$$

现在考虑任意一个元素 i 在元素 j 之前的状态，譬如说（…，i，i_1，…，i_k，j，…）。利用（$*$）逐次移位，我们有

$$\pi(\cdots,i,i_1,\cdots,i_k,j,\cdots)=\left(\frac{P_i}{P_j}\right)^{k+1}\pi(\cdots,i,i_1,\cdots,i_k,i,\cdots)$$

当 $P_j>P_i$ 时上式蕴含了

$$\pi(\cdots,i,i_1,\cdots,i_k,j,\cdots)<\frac{P_i}{P_j}\pi(\cdots,j,i_1,\cdots,i_k,i,\cdots)$$

令 $\alpha(i,j)=P\{e_i$ 在 e_j 之前$\}$，对一切 i 在 j 之前的状态求和，利用上式

可见

$$\alpha(i,j)<\frac{P_i}{P_j}\alpha(j,i)$$

由于 $\alpha(i,j)=1-\alpha(j,i)$，由此得

$$\alpha(j,i)>\frac{P_j}{P_j+P_i}$$

31. (a) 是。

(b) 处于状态 j 的时间的比例 $=\pi_j/\sum_{i=0}^{N}\pi_i$，$0\leqslant j\leqslant N$。

(c) 把到达 i 设想为更新，

$\pi_i(N)=(E[Y$ 链在相继到达 i 之间的转移次数$])^{-1}$

$\pi_j(N)=\dfrac{E[Y\text{ 链在相继到达 } i \text{ 之间到达 } j \text{ 的次数}]}{E[Y\text{ 链在相继到达 } i \text{ 之间的转移次数}]}$

$=\dfrac{E[X\text{ 链在相继到达 } i \text{ 之间到达 } j \text{ 的次数}]}{1/\pi_i(N)}$

(d) 对于对称随机游动，Y 链的转移概率是

$$P_{i,i+1}=\frac{1}{2}=P_{i,i-1},\quad i=1,\cdots,N-1$$

$$P_{00}=\frac{1}{2}=P_{01},\qquad P_{NN}=\frac{1}{2}=P_{N,N-1}$$

这转移概率矩阵是双随机的，因此

$$\pi_i(N)=\frac{1}{N+1},\quad i=0,1,\cdots,N$$

从而由 (b)

$E[X$ 链在相继到达 i 之间花在 j 上的时间$]=1$

(e) 利用定理4.7.1。

第五章

3. (a) 以 $N(t)$ 记到 t 为止的转移次数。容易证明这时

$$P\{N(t)\geqslant n\}\leqslant\sum_{j=n}^{\infty}e^{-Mt}\frac{(Mt)j}{j\,!}$$

从而 $P\{N(t)<\infty\}=1$。

5. 与其考虑一个尤尔过程，其 $X(0)=i$，不如设想有 i 个独立的尤尔过程，其 $X(0)=1$。对这 i 个群体的每一个在 t 时的总量取条件可见，k 个出生的条件分布是 k 个独立随机变量的分布，其中每个变量的分布由 (5.3.2) 给

出。

8. 我们从证明在时间 t 中有2个或更多个转移的概率是 $o(t)$ 开始。对下一个到达的状态取条件给出

$$P\{到时刻 t 有\geqslant 2个转移 | X_0=i\} = \sum_j P_{ij}P\{T_i+T_j\leqslant t\}$$

其中 T_i 与 T_j 是独立的参数分别为 v_i 与 v_j 的指数变量；它们分别表示离开 i 与 j 的时间。（注意到 T_i 与 j 是离 i 后到达的状态这一信息是独立的。）因此，

$$\begin{aligned}
&\lim_{t\to 0}P\{到时刻 t 有\geqslant 2个转移 | X_0=i\}/t \\
&\quad=\lim_{t\to 0}\sum_j P_{ij}P\{T_i+T_j\leqslant t\}/t \\
&\quad\leqslant\lim_{t\to 0}\Big[\sum_{j\leqslant M}P_{ij}P\{T_i+T_j\leqslant t\}/t+\sum_{j>M}P_{ij}P\{T_i\leqslant t\}/t\Big] \\
&\quad=\lim_{t\to 0}\left[\sum_{j\leqslant M}P_{ij}P\{T_i+T_j\leqslant t\}/t+\frac{1-e^{-v_it}}{t}\Big(1-\sum_{j\leqslant M}P_{ij}\Big)\right]
\end{aligned}$$

现在

$$P\{T_i-T_j\leqslant t\}\leqslant P\{T+T'\leqslant t\}=P\{N(t)\geqslant 2\}=o(t)$$

其中 T 与 T' 是独立的参数为 $v=\max(v_i,v_j)$ 的指数变量，$N(t)$ 是参数为 v 的泊松过程，而 $t\to 0$时

$$\frac{1-e^{-v_it}}{t}\to v_i$$

由此可知，对一切 M

$$\lim_{t\to 0}P\{到时刻 t 有\geqslant 2个转移 | X_0=i\}/t\leqslant v_i\Big(1-\sum_{j\leqslant M}P_{ij}\Big)$$

令 $M\to\infty$ 即得欲要的结果。

现在，

$$\begin{aligned}
P_{ii}(t)&=P\{X(t)=i | X(0)=i\} \\
&=P\{X(t)=i,到时间 t 未有转移 | X(0)=i\} \\
&\quad+P\{X(t)=i,到时间 t 至少有2个转移 | X(0)=i\} \\
&=e^{-v_it}+o(t)
\end{aligned}$$

类似地对 $i\neq j$

$$\begin{aligned}
P_{ij}(t)&=P\{第一个到达的状态是 j,转移时间\leqslant t | X(0)=i\} \\
&\quad+P\{X(t)=j,第一个到达的状态\neq j | X(0)=i\}
\end{aligned}$$

因此

$$P_{ij}(t)-P_{ij}(1-e^{-v_it})\leqslant P\{\text{到时间}\,t\,\text{有}\geqslant 2\text{个转移}\mid X(0)=i\}=o(t)$$

或

$$P_{ij}(t)=v_iP_{ij}t+o(t)$$

16. 以 q^x 及 P^x 记过程$\{X(t)\}$的转移率及平稳分布，类似地，对过程$\{Y(t)\}$记为 q^y,P^y。对链$\{(X(t),Y(t)),t\geqslant 0\}$有

$$q_{(i,j),(i',j)}=q^x_{i,i'}$$

$$q_{(i,j),(i,j')}=q^y_{j,j'}$$

我们说极限概率是

$$P_{i,j}=P^x_iP^y_j$$

为验证这点且同时证明时间可逆性，我们需要做的就是验证可逆性方程。现在

$$\begin{aligned}P_{i,j}q_{(i,j),(i',j)}&=P^x_iP^y_jq^x_{ii'}\\&=P^x_{i'}P^y_jq^x_{i',i}\quad(\text{由}\{X(t)\}\text{的时间可逆性})\\&=P_{i',j}q_{(i',j),(i,j)}\end{aligned}$$

由于对从(i,j)到(i,j')的转移的验证是类似的，故结论得证。

27. (a) $P_i/\sum_{j\in B}P_j$

(b) $$\begin{aligned}&P\{X(t)=i\mid X(t)\in B,X(t^-)\in G\}\\&=\frac{P\{X(t)=i,X(t^-)\in G\}}{P\{X(t)\in B,X(t^-)\in G\}}\\&=\frac{\sum_{j\in G}P\{X(t^-)=j\}P\{X(t)=i\mid X(t^-)=j\}}{\sum_{j\in G}P\{X(t^-)=j\}P\{X(t)\in B\mid X(t^-)=j\}}\\&=\sum_{j\in G}P_jq_{ji}\Big/\sum_{j\in G}\sum_{k\in B}P_jq_{jk}\end{aligned}$$

(c) 以 T 记离开状态 i 的时间，T' 记离开 i 之后直到进入 G 的附加时间。利用 T 与 T' 的独立性，对 i 之后到达的状态取条件即得

$$\begin{aligned}\tilde{F}_i(s)&=E[e^{-s(T+T')}]\\&=E[e^{-sT}]E[e^{-sT'}]\\&=\frac{v_i}{v_i+s}\sum_jE[e^{-sT'}\mid\text{下一个状态为}\,j]P_{ij}\end{aligned}$$

现在

$$E\,[e^{-sT'}\mid\text{下一个状态为}\,j]=\begin{cases}1, & \text{若}\ j\in G\\ \tilde{F}_j(s), & \text{若}\ j\in B\end{cases}$$

这就证得(c)。

(d) 在任意的时间 t 之中，从 G 到 B 的转移数与从 B 到 G 的转移数相差不超过1。因此长时间之后从 G 到 B 的转移率((d)的左边)必须等于从 B 到 G 的转移率((d)的右边)。

(e) 由(c)得

$$(s+v_i)\widetilde{F}_i(s)=\sum_{j\in B}\widetilde{F}_j(s)\,q_{ij}+\sum_{j\in G}q_{ij}$$

乘以 P_i 并对一切 $i\in B$ 相加得

$$\begin{aligned}\sum_{i\in B}P_i(s+v_i)\widetilde{F}_i(s)&=\sum_{i\in B}\sum_{j\in B}P_i\widetilde{F}_j(s)q_{ij}+\sum_{i\in B}\sum_{j\in G}P_iq_{ij}\\&=\sum_{j\in B}\widetilde{F}_j(s)\sum_{i\in B}P_iq_{ij}+\sum_{i\in G}\sum_{j\in B}P_iq_{ij}\\&=\sum_{j\in B}\widetilde{F}_j(s)[v_jP_j-\sum_{i\in G}P_iq_{ij}]+\sum_{i\in G}\sum_{j\in B}P_iq_{ij}\end{aligned}$$

其中最后的等式得自

$$v_jP_j=\sum_iP_iq_{ij},$$

而最后第二个等式得自(d)。由此可见

$$s\sum_{i\in B}P_i\widetilde{F}_i(s)=\sum_{i\in G}\sum_{j\in B}P_iq_{ij}(1-\widetilde{F}_j(s))$$

(f) $$E[e^{-sT_v}]=\frac{\sum_{i\in B}\widetilde{F}_i(s)\sum_{j\in G}P_jq_{ji}}{\sum_{j\in G}\sum_{k\in B}P_jq_{jk}}\qquad(\text{由(b)})$$

(g) 由(f)，利用(e)即得

$$s\sum_{i\in B}P_i\widetilde{F}_i(s)=\Big(\sum_{i\in G}\sum_{j\in B}P_iq_{ij}\Big)(1-E[e^{-sT_v}])$$

除以 s，再令 $s\to0$ 得

$$\sum_{i\in B}P_i=\sum_{i\in G}\sum_{j\in B}P_iq_{ij}E\,[T_v]$$

(h) $$E\,[e^{-sT_x}]=\frac{\sum_{i\in B}P_i\widetilde{F}_i\,(s)}{\sum_{j\in B}P_j}\qquad(\text{由(a)})$$

$$=\frac{\sum_{i\in G}\sum_{j\in B}P_iq_{ij}\,(1-\widetilde{F}_j(s))}{s\sum_{j\in B}P_j}\qquad(\text{由 (e)})$$

$$=\frac{\sum_{i\in G}\sum_{j\in B}P_iq_{ij}\ (1-E[e^{-ST_v}])}{s\sum_{j\in B}P_j} \qquad (\text{由 (f)})$$

$$=\frac{1-E\ [e^{-ST_v}]}{sE\ [T_v]} \qquad (\text{由}(g))$$

(i) 若 $P\{T_x\leqslant t\}$ 如所给定的，则

$$E[e^{-sT_x}]=\int_0^\infty e^{-st}P\{T_v>t\}dt/E[T_v]$$

$$=\int_0^\infty e^{-st}\int_t^\infty dF_{T_v}(y)dt/E[T_v]$$

$$=\int_0^\infty\int_0^y e^{-st}dtdF_{T_v}(y)/E[T_v]$$

$$=\frac{1-E[e^{-sT_v}]}{sE[T_v]}$$

因此，由(h)可知假定的分布产生正确的拉普拉斯变换。由变换与分布间的一一对应，结论得证。

(j) $E[T_x]=\int_0^\infty tdF_{T_x}(t)$

$$=\int_0^\infty t\int_t^\infty dF_{T_v}(y)dt/E[T_v]$$

$$=\int_0^\infty\int_0^y tdtdF_{T_v}(y)/E[T_v]$$

$$=E[T_v^2]/2E[T_v]$$

由于 $\mathrm{Var}(T_v)\geqslant 0$，$E[T_v^2]\geqslant(E[T_v])^2$

28. $\{R(t),t\geqslant 0\}$ 是一个两状态的马尔可夫链，它以指数率 $\lambda_1q(\lambda_2p)$ 离开状态 1(2) 进入状态 2(1)。由于 $P\{R(0)=1\}=p$，从例5.8(a)（或5.4(a)）得出

$$P\{R(t)=1\}=pe^{-\bar{\lambda}t}+(1-e^{-\bar{\lambda}t})\lambda_2p/\bar{\lambda}$$

其中 $\bar{\lambda}=\lambda_1q+\lambda_2p$。因此

$$(*)\qquad \int_0^t P\{R(s)=1\}ds=\frac{pq(\lambda_1-\lambda_2)}{\bar{\lambda}^2}(1-e^{-\bar{\lambda}t})+\frac{\lambda_2pt}{\bar{\lambda}}$$

为了证明(b)，令 $\Lambda(t)=\lambda_{R(t)}$，且对 $\varepsilon>0$ 写

$$N(t)=\sum_{n=1}^{t/\varepsilon}[N(n\varepsilon)-N((n-1)\varepsilon)]+o(\varepsilon)$$

现在

$$E[N(n\varepsilon)-N(n-1)\varepsilon|\Lambda((n-1)\varepsilon)]=\Lambda((n-1)\varepsilon)\varepsilon+o(\varepsilon)$$

因此

$$E[N(n\varepsilon)-N((n-1)\varepsilon)]=\varepsilon E[\Lambda((n-1)\varepsilon)]+o(\varepsilon)$$

由此可得

$$E[N(t)]=E\left[\sum_{n=1}^{t/\varepsilon}\varepsilon\Lambda(n-1)\varepsilon)+\frac{to(\varepsilon)}{\varepsilon}\right]$$

现在令 $\varepsilon\to0$得

$$\begin{aligned}E[N(t)]&=E[\int_0^t\Lambda(s)ds]\\&=\sum_{i=1}^{2}\lambda_i\int_0^t P\{R(s)=i\}ds\\&=\frac{pq(\lambda_1-\lambda_2)^2}{\bar{\lambda}^2}(1-e^{-\bar{\lambda}t})+\frac{\lambda_1\lambda_2 t}{\bar{\lambda}}\end{aligned}$$

这里最后的等式得自(*)。(b)的另一种论证(多少更直观些)如下:

$$P\{\text{在}(s,s+h)\text{ 中有更新}\}=hE[\Lambda(s)]+o(h)$$

所以,

$$E[\Lambda(s)]=P\{(s,s+h)\text{ 中有更新}\}/h+o(h)/h$$

令 $h\to0$给出

$$E[\Lambda(s)]=m'(s)$$

所以,

$$m(t)=\int_0^t E[\Lambda(s)]ds$$

第六章

3. 由(6.1.4)得出,从某个给定的时间开始,在布朗运动的值经时间 t_2-t_1改变 $B-A$ 的条件下,过程经时间 $s-t_1$所改变的值的条件分布是正态的,其均值为$(B-A)(s-t_1)/(t_2-t_1)$,方差为$(s-t_1)(t_2-s)/(t_2-t_1)$,$t_1<s<t_2$。因此,给定 $X(t_1)=A,X(t_2)=B,X(s)$是正态的,其均值与方差为

$$E[X(s)|X(t_1)=A,X(t_2)=B]=A+(B-A)\frac{s-t_1}{t_2-t_1}$$

$$\mathrm{Var}(X(s)|X(t_1)=A,X(t_2)=B)=\frac{(s-t_1)(t_2-s)}{t_2-t_1}$$

4. 对 $s\leqslant t$,

$$\mathrm{Cov}(X(s),X(t))=(s+1)(t+1)\mathrm{Cov}\left(Z\left(\frac{t}{t+1}\right),Z\left(\frac{s}{s+1}\right)\right)$$

由于$\{Z(t)\}$与$\{W(t)-tW(1)\}$同分布，其中$\{W(t)\}$是布朗运动，可见

$$\begin{aligned}\operatorname{Cov}(X(s),X(t)) &= (s+1)(t+1)\Bigg[\operatorname{Cov}\left(W\left(\frac{t}{1+t}\right),W\left(\frac{s}{1+s}\right)\right)\\ &\quad-\frac{t}{t+1}\operatorname{Cov}\left(W(1),W\left(\frac{s}{1+s}\right)\right)-\frac{s}{s+1}\operatorname{Cov}\left(W\left(\frac{t}{t+1}\right),W(1)\right)\\ &\quad+\frac{st}{(s+1)(t+1)}\operatorname{Cov}(W(1),W(1))\Bigg]\\ &=s(t+1)-st-st+st\\ &=s \qquad (s\leqslant t)\end{aligned}$$

由于易见$\{X(t)\}$是正态过程，$E[X(t)]=0$，故得它是布朗运动。

6. $\mu=\lambda,\quad c=\sqrt{1/2\lambda}$

7. 三个全有相同的密度。

14. 1/6

15. $f(x)=\dfrac{A-x}{\mu}+(B+A)\dfrac{e^{-2\mu A}-e^{-2\mu x}}{\mu(e^{2\mu B}-e^{-2\mu A})}$

16. (b) $E[T_x|X(h)]=E[h+T_{x-X(h)}]+o(h)$

$$=h+\frac{x-X(h)}{\mu}+o(h)\quad（因为 E[T_x]=\frac{x}{\mu}）$$

$$\begin{aligned}\operatorname{Var}(T_x|X(h)) &= \operatorname{Var}(h+T_{x-X(h)})+o(h)\\ &= g(x-X(h))+o(h)\end{aligned}$$

因此由条件方差公式

$$\begin{aligned}g(x)&=\operatorname{Var}(E[T_x|X(h)])+E[\operatorname{Var}(T_x|X(h))]\\ &=\frac{h}{\mu^2}+E[g(x-X(h))]+o(h)\\ &=\frac{h}{\mu^2}+E\left[g(x)-X(h)g'(x)+\frac{X^2(h)}{2}g''(x)+\cdots\right]+o(h)\\ &=\frac{h}{\mu^2}+g(x)-\mu hg'(x)+\frac{h}{2}g''(x)+o(h)\end{aligned}$$

除以 h，令 $h\to 0$即得

$$0=-\mu g'(x)+g''(x)/2+1/\mu^2$$

(c) $$\begin{aligned}\operatorname{Var}(T_{x+y})&=\operatorname{Var}(T_x+T_{x+y}-T_x)\\ &=\operatorname{Var}(T_x)+\operatorname{Var}(T_{x+y}-T_x)\qquad（由独立性）\\ &=\operatorname{Var}(T_x)+\operatorname{Var}(T_y)\end{aligned}$$

所以

$$g(x+y)=g(x)+g(y)$$

这蕴含

$$g(x) = cx$$

(d)由(c)可见

$$g(x) = cx$$

由(b)这蕴含

$$g(x) = x/\mu^3$$

17. $\min(1, 5x/4)$

第七章

4. $0 \leqslant \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = n\mathrm{Var}(X_1) + n(n-1)\mathrm{Cov}(X_1, X_2)$

$\therefore$ 对一切 n, $\mathrm{Var}(X_1) + (n-1)\mathrm{Cov}(X_1, X_2) \geqslant 0 \Rightarrow \mathrm{Cov}(X_1, X_2) \geqslant 0$

作为有限情形的反例，令 X_1 为零均值正态变量，$X_2 = -X_1$。

5. $\mathrm{Var} Z_n = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j}\sum \mathrm{Cov}(X_i, X_j)$$

但对 $i < j$,

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= E[X_i X_j] \\
&= E[(Z_i - Z_{i-1})(Z_j - Z_{j-1})] \\
&= E[E[(Z_i - Z_{i-1})(Z_j - Z_{j-1}) \mid Z_1, \cdots, Z_i]] \\
&= E[(Z_i - Z_{i-1})(E[Z_j \mid Z_1, \cdots, Z_i] - E[Z_{j-1} \mid Z_1, \cdots, Z_i])] \\
&= E[(Z_i - Z_{i-1})(Z_i - Z_i)] \\
&= 0
\end{aligned}$$

8. $E[S_n^2 \mid S_1^2, \cdots, S_{n-1}^2]$

$$\begin{aligned}
&= E[(S_{n-1} + X_n)^2 \mid S_1^2, \cdots, S_{n-1}^2] \\
&= E[S_{n-1}^2 \mid S_1^2, \cdots, S_{n-1}^2] + 2E[X_n S_{n-1} \mid S_1^2, \cdots, S_{n-1}^2] \\
&\quad + E[X_n^2 \mid S_1^2, \cdots, S_{n-1}^2] \\
&= S_{n-1}^2 + 2E[X_n]E[S_{n-1} \mid S_1^2, \cdots, S_{n-1}^2] + E[X_n^2] \\
&= S_{n-1}^2 + \sigma^2
\end{aligned}$$

因此，

$$E[S_n^2 - n\sigma^2 \mid S_1^2, \cdots, S_{n-1}^2] = S_{n-1}^2 - (n-1)\sigma^2$$

13. 利用 $f(x) = e^{\theta x}$ 是凸函数的事实，然后由延森不等式

$$1 = E[e^{\theta X}] \geqslant e^{\theta E[X]}$$

由于 $E[X]<0$，上式蕴含 $\theta>0$。

第八章

8. (a) 利用表示式

$$X = \sum_{i=1}^{n} X_i$$

其中 $X_1, \cdots, X_n$ 相互独立，且

$$P\{X_i = 1\} = 1 - P\{X_i = 0\} = p, \qquad i = 1, \cdots, n$$

由于 X_i 是离散 IFR，由有关卷积的结果，X 也是。

(b) 设 X_n 是参数为 n, p_n 的二项分布变量，其中 $np_n = \lambda$。由 IFR 性在极限下保持不变，而 X_n 收敛于均值为 λ 的泊松随机变量，结论得证。

(c) 设 $X_i, i \leqslant 1$，为独立的几何分布随机变量，即

$$P\{X_i = k\} = p(1-p)^{k-1}, k \geqslant 1$$

由于

$$P\{X_i = k \mid X_i \geqslant k\} = p$$

故得 X_i 是离散 IFR，$i \geqslant 1$。因此 $\sum_{i=1}^{r} X_i$ 是 IFR，故得结论。

12. 以 T_i 记马尔可夫链中从 i 到达的下一个状态。题设给出 $T_i \underset{\text{st}}{\leqslant} T_{i+1}, i \geqslant 1$。因此由 $\sum_j P_{ij} f(s) = E[f(T_i)]$ 得(a)。为证(b)对 n 用归纳法如下：以 P_i 与 E_i 表示在条件 $X_0 = i$ 之下的概率与期望，我们有

$$\begin{aligned} P_i\{X_n \geqslant k\} &= E_i[P_i\{X_n \geqslant k \mid X_1\}] \\ &= E_i[P_{X_1}\{X_{n-1} \geqslant k\}] \end{aligned}$$

归纳法假设说 $P_i\{X_{n-1} \geqslant k\}$ 关于 i 递增，因此 $P_{X_1}\{X_{n-1} \geqslant k\}$ 是 X_1 的递增函数。由(a)，X_1 关于初始状态 i 随机递增，所以对任一递增函数 $g(X_1)$，特别对 $g(X_1) = P_{X_1}\{X_{n-1} \geqslant k\}$，我们有

$$E_i[g(X_1)] \uparrow i$$

17. 假设 Y 与 Z 独立，$Y \sim G, \min(Y, Z) \sim F$。利用两个独立随机变量的最小值的失效率函数等于它们的失效率函数之和，给出

$$\lambda_G(t) = \lambda_Y(t) \leqslant \lambda_Y(t) + \lambda_Z(t) = \lambda_F(t)$$

为证另一方向的结论，设 $\lambda_F(t) \geqslant \lambda_G(t)$。设 Y 有分布 F，定义 Z 与 Y 独立，

且有失效率函数

$$\lambda_Z(t) = \lambda_F(t) - \lambda_G(t)$$

随机变量 Y,Z 满足所要求的条件。

26. 令 $n=2$，则

$$P\{X_1 + X_2 \geqslant 2\} = P_1P_2 \leqslant \left(\frac{P_1 + P_2}{2}\right)^2 = P\{\text{Bin}(2,\bar{p}) \geqslant 2\}$$

$$\sum_{i=1}^{2} P\{X_1 + X_2 \geqslant i\} = P_1P_2 + P_1 + P_2 - P_1P_2 = \sum_{i=1}^{2} P\{\text{Bin}(2,\bar{p}) \geqslant i$$

这证明了 $$X_1+X_2 \underset{\text{v}}{\leqslant} \text{Bin}(2,\bar{p})$$

现在考虑 n 的情形，且设 $P_1 \leqslant P_2 \leqslant \cdots \leqslant P_n$。设 f 为递增凸函数，用 $n=2$ 时的结论可见

$$E\left[f\left(\sum_{i=1}^{n} X_i\right) | X_2,\cdots,X_{n-1}\right] \leqslant E\left[f\left(\bar{X}_1 + \bar{X}_n + \sum_{i=2}^{n-1} X_i\right) | X_2,\cdots,X_{n-1}\right]$$

其中，$\bar{X}_1,\bar{X}_n$ 与别的变量都独立，且是贝努里随机变量，

$$P\{\bar{X}_1 = 1\} = P\{\bar{X}_n = 1\} = \frac{P_1 + P_n}{2}$$

在上面的不等式中取期望证得

$$\sum_{i=1}^{n} X_i \underset{\text{v}}{\leqslant} \bar{X}_1 + \bar{X}_n + \sum_{i=2}^{n-1} X_i$$

重复这段论证(从而连续地证明了，现在的诸 X 的和比把 P 最大的 X 与 P 最小的 X 换成两个有它们的 P 的平均值的 X 后的和较少变)并取极限得

$$\sum_{i=1}^{n} X_i \underset{\text{v}}{\leqslant} \text{Bin}(n,\bar{p})$$

现在写

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i + \sum_{i=n+1}^{n-m} X_i$$

其中 $X_i \equiv 0, n+1 \leqslant i \leqslant n+m$。由已证结论

$$\sum_{i=1}^{n} X_i \underset{\text{v}}{\leqslant} \text{Bin}\left(n + m, \sum_{i=1}^{n} \frac{P_i}{n + m}\right)$$

令 $m \to \infty$ 得

$$\sum_{i=1}^{n} X_i \underset{\text{v}}{\leqslant} \text{Poisson}\left(\sum_{i=1}^{n} P_i\right) \qquad \text{（参数为} \sum_{i=1}^{n} P_i \text{ 的泊松变量）}$$

注：我们假定变动性对取极限仍然保持。

29. 设 f 为凸函数。$f(cX)$关于 X 的泰勒级数展开给出

$$f(cX)=f(X)+f'(X)(c-1)X+f''(Z)(cX-Z)^2/2$$

其中 $X<Z<cX$。取期望并利用 f 的凸性给出

$$E[f(cX)]\geqslant E[f(X)]+(c-1)E[Xf'(X)]$$

现在对 $X\geqslant 0$，函数 $g(X)=X$ 与 $g(X)=f'(X)$关于 X 都递增(后者由于 f 的凸性)，所以由第七章的命题7.1.5

$$E[Xf'(X)]\geqslant E[X]E[f'(X)]=0$$

这就证明了 $E[f(cX)]\geqslant E[f(X)]$。

名词索引

E

F

J

K

L

M

N

O

P

T